Classificação de Sinais de ECG com Técnicas Explicáveis de Inteligência Artificial

Hugo Fidalgo Oliveira Martins

Dissertação apresentada à Escola de Tecnologia e Gestão no âmbito do Mestrado em Engenharia Eletrotécnica e de Computadores

Orientador:

Prof. Dr. João Paulo Teixeira

Bragança

Outubro, 2025

Agradecimentos

A presente dissertação encerra mais um ciclo do meu percurso académico, o qual apenas foi possível graças ao contributo e apoio de diversas pessoas às quais agradeço.

Em primeiro lugar, agradeço ao meu orientador, Professor Doutor João Paulo Teixeira, pela orientação científica, disponibilidade e confiança depositada ao longo de todo o processo. As suas sugestões e acompanhamento foram fundamentais para a concretização desta dissertação.

Em segundo lugar, gostaria de agradecer ao Instituto Politécnico de Bragança, pela oportunidade de realizar mais um objetivo académico, pelo ambiente agradável e motivador, que se repercutiram na qualidade de ensino e na disponibilização de todos os recursos e condições necessárias para a concretização de metas e preparação para o mercado profissional.

Por último, à minha família pela paciência e compreensão demonstrados ao longo destes anos de estudo e à minha namorada Mariana, por todo o carinho, paciência e apoio incondicional, que foram fundamentais para que este trabalho se tornasse possível.

Abstract

Cardiovascular diseases are among the main causes of premature mortality, with atrial flutter representing a clinically relevant arrhythmia due to its association with stroke and heart failure. The eletrocardiogram (ECG) is the most suitable diagnostic method for evaluating chardiac rhytms. Automatic ECG interpretations have attempted to improve clinical practice. However, the lack of interpretability of existing models has limited their acceptance.

This dissertation presents a framework for atrial flutter classification using raw 12-lead ECG signals from PTB-XL Database, Georgia 12-Lead ECG Challenge Database and Large 12-Lead ECG Database for Arrhythmia Study. A hybrid deep learning model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks was developed and trained under a 10-fold cross-validation scheme. To ensure model transparency, explainable artificial intelligence (XAI) mehods were applied: *Shapley Additive Explanations* (SHAP) was used to quantify the contribution of each lead, and *Local Interpretable Model-Agnostic Explanations* (LIME) was employed to highlight the most informative temporal segments at the patient level.

The results demonstrate that the proposed approach achieves competitive performance while improving interpretability, thus contributing to more reliable and clinically meaningful applications of artificial intelligence in cardiology.

Keywords: Eletrocardiogram (ECG), Atrial Flutter (AFL), Deep Learning (DL), Explainable AI (XAI)

Resumo

As doenças cardiovasculares estão entre as principais causas de mortalidade prematura, sendo o Flutter Auricular uma arritmia clinicamente relevante devido à sua associação com acidente vascular cerebral e insuficiência cardíaca. O eletrocardiograma (ECG) constitui uma ferramenta de diagnóstico fundamental, e os recentes avanços em aprendizagem profunda têm melhorado a classificação automática dos ritmos cardíacos. No entanto, a dificuldade de interpretação dos modelos existentes limita a sua aceitação na prática clínica.

Na presente dissertação apresenta-se uma abordagem para a classificação do Flutter Auricular utilizando sinais brutos de ECG de 12 derivações, provenientes das bases de dados PTB-XL Database, Georgia 12-Lead ECG Challenge Database e Large 12-Lead ECG Database for Arrhythmia Study. Foi desenvolvido e treinado um modelo híbrido de aprendizagem profunda, que combina Redes Neuronais Convolucionais (CNN) e Redes Long Short-Term Memory (LSTM), recorrendo a um esquema de validação cruzada de 10 *folds*. Para garantir a transparência do modelo, foram aplicadas técnicas de inteligência artificial explicável (XAI): o *Shapley Additive Explanations* (SHAP) foi utilizado para quantificar a contribuição de cada derivação, e o *Local Interpretable Model-Agnostic Explanations* (LIME) foi aplicado para destacar os segmentos temporais mais informativos ao nível do paciente.

Os resultados demonstram que a abordagem proposta apresenta um desempenho competitivo, ao mesmo tempo que melhora a interpretabilidade, contribuindo assim para aplicações de inteligência artificial mais fiáveis e clinicamente relevantes em cardiologia.

Palavras-chave: Eletrocardiograma (ECG), Flutter Auricular (AFL), Aprendizagem Profunda (DL), Inteligência Artificial Explicável (XAI)

Índice

Lista de Figuras

Abreviaturas

AF Fibrilhação auricular.

AFL Flutter auricular.

CNN Convolutional neural network.

DCV Doenças cardiovasculares.

DL Deep learning.

DNN Deep neural network.

ECG Eletrocardiograma.

GRAD-CAM Gradient-weighted Class Activation Mapping.

IPB Instituto Politécnico de Bragança.

LIME Local Interpretable Model-Agnostic Explanations.

LSTM Long-short term memory network.

ML Machine learning.

OMS Organização Mundial de Saúde.

RNN Recurrent neural network.

SHAP Shapley Additive Explanations.

XAI Explainable artificial intelligence.

Capítulo 1

Introdução

De acordo com a Organização Mundial de Saúde (2021) (OMS), as doenças cardiovasculares (DCV) constituem umas das principais causas de mortalidade a nível global, tendo sido responsáveis por cerca de 17.9 milhões de mortes em 2019, número que se prevê que possa atingir os 24 milhões em 2030 (Yıldırım et al. 2018). As DCV incluem patologias como enfarte agudo do miocárdio, acidente vascular cerebral e arritmias cardíacas, estando, por isso, associadas a morbilidade elevada, perda de qualidade de vida e custos socioeconómicos significativos (Saini e Gupta 2022).

Num coração saudável, os impulsos elétricos gerados no nódulo sinoauricular seguem um padrão regular, originando sinais elétricos estáveis (Teixeira e Lopes 2011; Saini e Gupta 2022). Contudo, perturbações neste sistema podem tornar os impulsos anómalos, resultando em batimentos cardíacos demasiado rápidos, lentos ou irregulares – condição conhecida como arritmia cardíaca (Apandi et al. 2018; Saini e Gupta 2022). Estima-se que entre 2% e 5% da população mundial possa sofrer de algum tipo de arritmia (Khurshid et al. 2018), sendo a fibrilhação auricular (AF) e o flutter auricular (AFL) comuns e associadas a risco aumentado de acidente vascular cerebral e insuficiência cardíaca (Faust et al. 2018; Borghi et al. 2021). O diagnóstico precoce destas condições é, por isso, fundamental para melhorar o prognóstico do paciente e reduzir os custos associados a tratamentos mais invasivos em fases avançadas da doença (Zeng et al. 2024).

O eletrocardiograma (ECG) é o método de diagnóstico mais utilizado para o registo

da atividade elétrica cardíaca (Apandi et al. 2018; Finocchiaro et al. 2020; Costa et al. 2021; Saini e Gupta 2022; Zeng et al. 2024). Este exame médico é tipicamente realizado através de um sistema de 12 derivações, obtidas a partir de 10 elétrodos aplicados na pele — três derivações dos membros, três derivações dos membros aumentadas e seis derivações precordiais (torácicas), formadas entre os elétrodos físicos e um elétrodo virtual conhecido como *terminal central de Wilson* (Yao et al. 2020). A Figura 1.1(a) ilustra o posicionamento dos elétrodos e a Figura 1.1(b) apresenta a disposição espacial dos elétrodos e os diferentes ângulos de captação dos sinais sobre a pele.

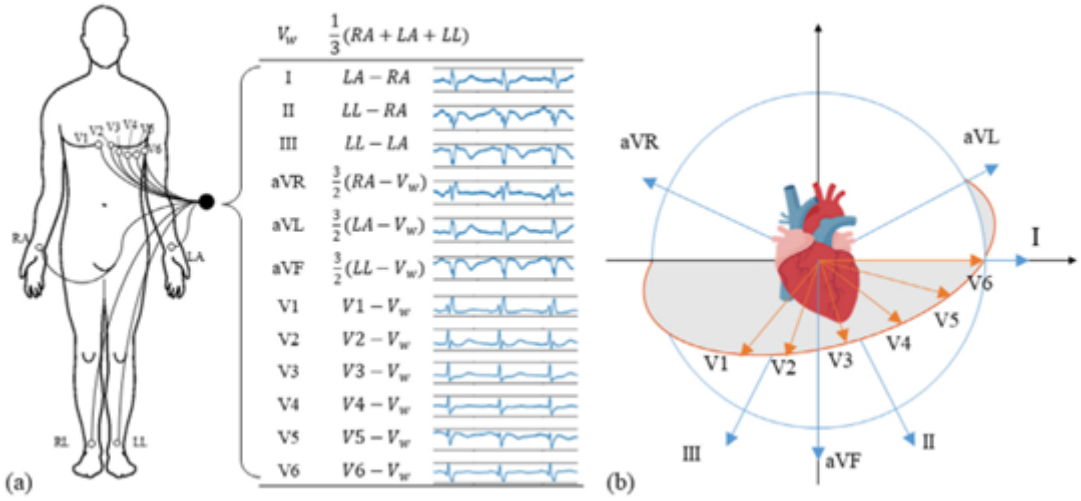


Figura 1.1: Representação da posição das 12-derivações e correspondente registo do Eletrocardiograma (a) e diferentes ângulos de captação dos sinais na pele (b) (retirado de Yao et al. (2020))

Entre as 12 derivações de Eletrocardiograma, a derivação II (LL-RA) é uma das mais utilizadas para a identificação de patologias, por evidenciar de forma clara as ondas P, o complexo QRS e a onda T ao longo do ciclo cardíaco (Luz et al. 2016; Yıldırım et al. 2018). As derivações precordiais, em especial V1 e V2, são relevantes para a deteção de anomalias relacionadas com o funcionamento dos ventrículos, dada a sua posição torácica (Luz et al. 2016).

Na Figura 1.2 apresenta-se o traçado típico de um ECG, composto por diferentes deflexões correspondentes aos fenómenos elétricos que ocorrem na superfície do coração,

designados por despolarização atrial (onda P), despolarização ventricular (complexo QRS) e repolarização ventricular (onda T). As arritmias podem provocar alterações significativas na morfologia e na frequência destas ondas, tornando possível a sua identificação através da análise do traçado (Luz et al. 2016; Yıldırım et al. 2018).

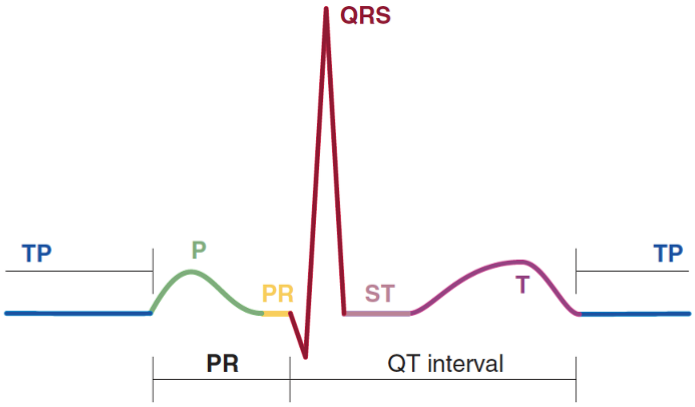


Figura 1.2: Traçado de um ECG normal (retirado de Ayano et al. (2022))

Existem diversos tipos de arritmias, cada um associado a um padrão de ECG característico, o que possibilita a sua identificação e classificação (Luz et al. 2016). De forma geral, as arritmias podem ser agrupadas em duas categorias principais: (i) arritmias morfológicas, resultantes de um único batimento cardíaco irregular, e (ii) arritmias rítmicas, caracterizadas por um conjunto de batimentos irregulares (Luz et al. 2016). Exemplos de arritmias rítmicas incluem a AF e o AFL, ambos visíveis no ECG por meio de padrões bem definidos (Luz et al. 2016).

O flutter auricular é uma arritmia que se caracteriza por contrações auriculares rápidas e regulares (Cosío, 2017). No traçado de ECG, o flutter corresponde a uma ondulação contínua e regular entre os complexos QRS (Cosío, 2017). Visualmente, este padrão distingue-se pela ausência de ondas P bem definidas, sendo substituídas por uma morfologia típica em "dente de serra", com uma frequência de 250 a 350 bpm, tal como mostra a Figura 1.3 (Teixeira e Lopes 2011)

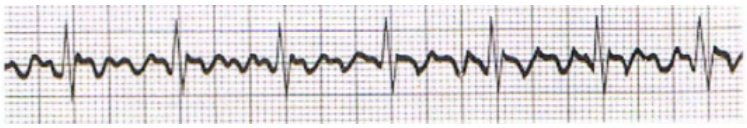


Figura 1.3: ECG típico de um flutter auricular (retirado de Teixeira e Lopes (2011))

Os algoritmos existentes para o reconhecimento automático de arritmias cardíacas em ECG baseiam-se, na sua maioria, na análise de características morfológicas de um único ou de poucos complexos QRS (Yıldırım et al. 2018). Esta abordagem, embora eficaz, é suscetível a erros devido à elevada variabilidade *beat-to-beat* entre indivíduos (Luz et al. 2016; Yıldırım et al. 2018). Adicionalmente, a interpretação visual do ECG é um processo demorado e sujeito a subjetividade (Sinha et al. 2022; Zeng et al. 2024). Surge, assim, a necessidade de métodos automáticos e robustos para a classificação de arritmias (Luz et al. 2016).

Neste contexto, técnicas de *Deep Learning* (DL) têm-se revelado promissoras, em particular arquiteturas como a *Convolutional Neural Network* (CNN) e a *Long Short-Term Memory Network* (LSTM), que têm apresentado resultados significativos na classificação de DCV (Acharya et al. 2017; Faust et al. 2018; Sreeja e Supriya 2023). A aplicação combinada destas técnicas com métodos de Inteligência Artificial Explicável (XAI), como SHAP e LIME, permite não só uma classificação mais precisa, mas também uma interpretação transparente das decisões do modelo – fator essencial para a aceitação clínica e conformidade com diretrizes éticas e regulamentares (Ardeti et al. 2023; Prakash et al. 2023; Zeng et al. 2024).

Partindo do exposto, a presente dissertação teve como objetivo a classificação de sinais ECG e a sua associação a patologias cardíacas, recorrendo a modelos de DL baseados em CNN e LSTM, complementados por métodos de XAI, de forma a proporcionar maior interpretabilidade e confiança nos resultados.

1.1 Contributos do Estudo

Dada a relevância e atualidade do tema em estudo, é esperado que as conclusões da presente investigação contribuam significativamente para a literatura científica na área da deteção de DCV recorrendo a métodos de XAI. Adicionalmente, espera-se que os resultados obtidos sejam úteis tanto para a comunidade médica como para a comunidade científica, facilitando o desenvolvimento de soluções mais precisas e transparentes para a análise automática de sinais ECG e potenciando a adoção de sistemas de apoio à decisão clínica baseados em inteligência artificial.

1.2 Estrutura do Estudo

No Capítulo 2 é apresentada uma revisão de literatura, onde são aprofundados conceitos essenciais ao estudo e analisados trabalhos previamente publicados na área. O Capítulo 3 apresenta detalhadamente a metodologia usada, nomeadamente a recolha dos sinais ECG, a criação e treino da rede neuronal e a análise de interpretabilidade com SHAP e LIME. De seguida, no Capítulo 4, é feita a análise dos resultados obtidos e, finalmente, no Capítulo 5 é feita uma discussão dos resultados comparando-os com trabalhos anteriores e são apresentadas as principais conclusões do estudo.

Capítulo 2

Revisão de Literatura

As técnicas de DL têm ganho destaque como ferramenta eficaz para analisar sinais de ECG e deteção atempada de DCV (Raza et al. 2022; Zeng et al. 2024). Este capítulo apresenta as principais metodologias de DL e XAI aplicadas a sinais de ECG, bem como estudos prévios que fundamentam a abordagem desta investigação.

2.1 Técnicas de Deep Learning (DL)

As técnicas de aprendizagem profunda - *Deep Learning* (DL), fazem parte da aprendizagem máquina - *Machine learning* (ML) (Yıldırım et al. 2018) e são caracterizadas pelo uso de arquiteturas hierárquicas compostas por múltiplas camadas, nas quais a informação é processada de forma sequencial e progressiva (LeCun et al. 2015; Goodfellow et al. 2016; Yıldırım et al. 2018).

Ao contrário de muitas abordagens clássicas de ML, que exigem a extração manual de características dos sinais, os modelos DL realizam essa etapa automaticamente através de camadas de entrada – *input layers* (Zeng et al. 2024). Estas camadas extraem as informações relevantes, permitindo que as camadas de saída (*output layers*) analisem e classifiquem padrões identificados (LeCun et al. 2015; Yıldırım et al. 2018; Wasimuddin et al. 2020; Anand et al. 2022).

De forma geral, as técnicas de DL podem ser classificadas em diferentes categorias,

consoante a abordagem de treino utilizada (Yıldırım et al. 2018). Entre os modelos discriminativos destacam-se a *Deep Neural Network* (DNN), a *Recurrent Neural Network* (RNN) e a *Convolutional Neural Network* (CNN). Já no grupo de modelos não supervisionados ou generativos, encontram-se, por exemplo, os *Autoencoders* regularizados e as *Deep Boltzman Machine* (Schmidhuber 2015; Yıldırım et al. 2018).

2.1.1 *Convolutional Neural Networks* (CNN)

Entre os modelos discriminativos de redes neuronais profundas, as CNN representam um subtipo de rede neuronal *feedforward* com estrutura hierárquica (Liu et al. 2021). Em vez de dependerem exclusivamente de camadas totalmente conectadas, como acontece nas redes neuronais tradicionais, as CNN utilizam filtros de aprendizagem que aplicam operações de convolução a sub-regiões da entrada, extraindo automaticamente características relevantes (Yıldırım et al. 2018; Liu et al. 2021).

Em termos de arquitetura, uma CNN é composta tipicamente pelos seguintes blocos principais (LeCun et al. 2015; Ping et al. 2020; Liu et al. 2021):

1. **Camadas convolucionais** – responsáveis pela extração de *features* a partir dos dados de entrada, recorrendo a *kernels* de convolução treináveis;

2. **Camadas de agrupamento** (*pooling layers*) – executam operações de redução de dimensionalidade (*down-sampling*), preservando a informação mais representativa e suprimindo redundâncias;

3. **Camadas totalmente conectadas** (*fully connected layers*) – onde as *features* extraídas são transformadas em representações lineares que alimentam a etapa final de classificação.

Desta forma, e de acordo com Liu et al. (2021), o funcionamento típico de uma camada convolucional pode ser descrito matematicamente pela seguinte Equação (2.1) para o j-ésimo mapa de características da l-ésima camada:

$$c_j^l = \theta\left(\sum_{i \in M_j} x_i^{l-1} * w_{ij}^l + b_j^l\right) \tag{2.1}$$

onde θ representa a função de ativação e M_j representa a conectividade entre c_j^l e os mapas de características da camada anterior. w_{ij}^l corresponde ao peso (ou *kernel*) para o j-ésimo mapa de características e o i-ésimo índice de filtro, e b_j^l é o viés correspondente.

De seguida, a camada de *pooling* reduz o tamanho das *features* através de *down-sampling*, de modo a selecionar os dados mais representativos da amostra. Para cada bloco de dados, é produzida uma única saída correspondente.

Após as camadas convolucionais e de *pooling*, as *features* de cada sub-região são achatadas (*flattened*) para formar um vetor unidimensional que serve como entrada para a camada totalmente conectada (*fully connected layer*). Nesta fase, os dados de entrada são mapeados para classes específicas. O processo de treino utiliza retropropagação (*back-propagation*), em conjunto com dados rotulados e uma taxa de aprendizagem (*learning rate*) definida, para otimizar os parâmetros da rede segundo a função de custo estabelecida.

Segundo a literatura, as CNN distinguem-se por apresentarem um número inferior de parâmetros a otimizar, o que se traduz em treinos mais simples e mais eficientes, com menor risco de sobreajuste (Yıldırım et al. 2018). Diversos estudos comprovam a sua eficácia na análise de sinais ECG. Por exemplo, Acharya et al. (2017) desenvolveram uma CNN com nove camadas para classificar cinco tipos de batimentos cardíacos em sinais ECG, atingindo uma precisão de 94.03% em dados originais e 93.47% em dados aumentados por técnicas de *data augmentation*.

De forma semelhante, Yıldırım et al. (2018) propuseram uma CNN unidimensional para a classificação automática de arritmias cardíacas na base de dados *MIT-BIH Arrhythmia*, obtendo uma precisão global de 91.33% na classificação de 17 tipos de arritmias, com tempo médio de 0,015 segundos por amostra, demonstrando potencial para

aplicações em tempo real.

Estudos recentes reforçam esta tendência: Chen et al. (2020) e Zeng et al. (2024) destacam que as CNN, ao aprenderem automaticamente representações hierárquicas dos sinais ECG através de padrões locais e dependências espaciais, tornam-se particularmente adequadas para a classificação de DCV.

2.1.2 *Long short-term memory* (LSTM)

As *Recurrent Neural Networks* (RNN) são um tipo de rede neuronal projetada para processar dados sequenciais, estabelecendo conexões recorrentes entre unidades de memória que permitem reter informação ao longo do tempo (Faust et al. 2018; Liu et al. 2021). Neste sentido, ao possibilitar a modelação de dependências temporais, as RNN são adequadas para tarefas que envolvam séries temporais, como os sinais ECG.

Contudo, as RNN convencionais apresentam limitações quando aplicadas a dependências de longo prazo (Liu et al. 2021). Desta forma, Hochreiter e Schmidhuber (1997) propuseram a arquitetura *Long Short-Term Memory* (LSTM), capaz de capturar dependências tanto de curto como de longo prazo em sequências temporais (Faust et al. 2018; Kłosowski et al. 2020; Zeng et al. 2024).

As LSTM introduzem um mecanismo de células de memória dotadas de portas de controlo (*gates*), que regulam seletivamente o fluxo de informação, garantindo que a informação útil seja armazenada e a redundante descartada (Liu et al. 2021).

Na Figura 2.1 apresenta-se de forma esquemática o funcionamento de uma LSTM.

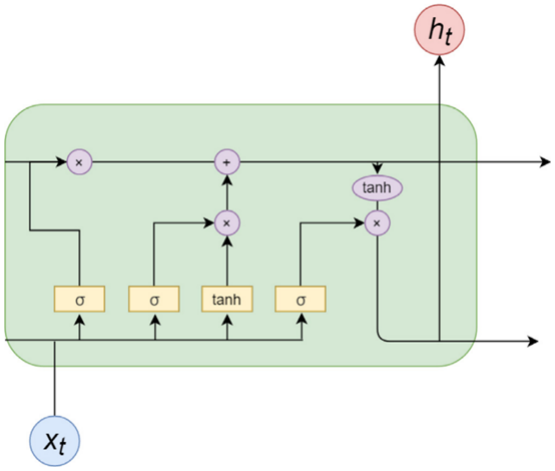


Figura 2.1: Estrutura de um bloco simples LSTM (imagem retirada de Liu et al. (2021))

O funcionamento de uma LSTM contém três portas principais (Ping et al. 2020; Liu et al. 2021):

- **Forget gate** – decide a informação a ser descartada da célula de memória. Assim, para uma entrada x_t e a saída do passo temporal anterior h_{t-1}, é dada pela equação (2.2):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.2}$$

- **Input gate** – regula a incorporação de nova informação, sendo dada pela equação (2.3):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2.3}$$

  onde f_t representa a *forget gate* e i_t a *input gate*. σ é a função de ativação sigmoide, que restringe os valores entre 0 e 1, controlando se a informação é armazenada.

  De seguida, é gerada uma nova memória $\tilde{c}_t$ (2.4), e a memória anterior c_{t-1} (2.5) é atualizada:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{2.4}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{2.5}$$

- **Output gate** – controla a informação que será transmitida como saída para o próximo estado oculto. A saída h_t é calculada pelas seguintes equações (2.6) e (2.7):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{2.6}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{2.7}$$

Tendo em conta a natureza tipicamente temporal dos sinais ECG, esta técnica é vastamente explorada na literatura (Cheng et al. 2021; Liu et al. 2021). Chang et al. (2018) treinaram uma rede LSTM para analisar variações temporais nos intervalos RR dos sinais ECG, alcançando uma precisão de 98.3% na deteção de fibrilhação auricular.

Kłosowski et al. (2020) transformaram séries temporais de ECG em imagens espectrais, com recurso à transformada de Fourier de curto prazo, e mostraram que esta transformação melhora o desempenho da LSTM.

De forma complementar, Karri e Annavarapu (2023) desenvolveram um sistema embutido para a deteção do complexo QRS e a subsequente classificação de arritmias. Para isso, utilizaram, para a deteção, a modulação *Delta-Sigma* e a transformada de *Wavelet* discreta e, para a classificação de arritmias, a rede neuronal LSTM. Com uma precisão de 99.64%, concluíram que o modelo LSTM é altamente eficiente.

2.1.3 Modelos Híbridos (CNN-LSTM)

Apesar do sucesso individual das CNN e das LSTM, diversos estudos mostram que a combinação das duas arquiteturas potencia resultados superiores na análise de sinais biomédicos, nomeadamente nos sinais ECG. Esta abordagem híbrida permite capturar

simultaneamente características espaciais extraídas das camadas convolucionais e as dependências temporais provenientes das unidades LSTM (Liu et al. 2021).

Estudos recentes comprovam a eficácia desta abordagem, como é o caso de Ping et al. (2020), que propuseram um modelo composto por oito camadas CNN, uma conexão *shortcut* e uma camada LSTM para a classificação de fibrilhação auricular (AF). Com recurso a *ten-fold cross-validation* e com segmentações do sinal de 5, 10 e 20 segundos, obtiveram F1-Score de, respetivamente, 84,89%, 89,55% e 85,64%. Estes resultados demonstram a robustez do modelo e o potencial para aplicação na prática médica.

De forma semelhante, Chen et al. (2020) desenvolveram um classificador híbrido CNN-LSTM para a classificação de seis tipos de arritmias, combinando segmentos de 10 segundos e respetivos intervalos RR. O modelo alcançou 99.32% de precisão na base de dados MIT-BIH Arrhythmia e uma precisão média de 97.15% em duas bases de dados independentes, evidenciando elevada capacidade de generalização.

Madan et al. (2022) adotaram uma estratégia alternativa, convertendo os sinais de ECG em imagens de escalograma bidimensionais e, com recurso a um modelo híbrido CNN-LSTM, atingiram aproximadamente 99% de precisão na deteção de arritmia cardíaca, insuficiência cardíaca congestiva e ritmo sinusal normal, reforçando o potencial da abordagem multimodal.

De acordo com a literatura, a combinação de arquiteturas CNN e LSTM mostra-se como uma solução promissora na classificação automática de sinais ECG, ao conjugar a robustez da extração de *features* hierárquicas com dependências de curto e longo prazo (Cheng et al. 2021; Liu et al. 2021).

2.1.4 Limitações das técnicas de Deep Learning (DL)

A literatura reforça os avanços na utilização de redes neuronais profundas na classificação de sinais ECG. Modelos como CNN, LSTM e o seu uso combinado têm demonstrado elevada precisão na deteção de DCV (Acharya et al. 2017; Faust et al. 2018; Sreeja e Supriya 2023). No entanto, subsistem limitações que condicionam a sua adoção plena

em contextos clínicos. Uma das principais limitações é a capacidade de generalização, isto é, modelos de DL tendem a ter dificuldade em lidar com a adaptação a novos sinais de ECG, provenientes de outras fontes que não as previamente conhecidas (Ribeiro et al. 2016; Singh e Sharma 2022; Abdullah et al. 2023). Este problema decorre, por exemplo, do desbalanceamento das bases de dados, da variabilidade interindividual e da sensibilidade dos modelos.

Outra limitação inerente aos modelos de DL está relacionada com a falta de transparência, uma vez que são recorrentemente classificadas como "*black boxes*" (Singh e Sharma 2022; Alamatsaz et al. 2024). Isto deve-se à dificuldade de interpretação dos processos internos de tomada de decisão (Singh e Sharma 2022; Alamatsaz et al. 2024). Tal opacidade gera ceticismo na comunidade médica, que precisa de explicações claras e fundamentadas para validar decisões automáticas (Ribeiro et al. 2016; Singh e Sharma 2022; Abdullah et al. 2023).

Desta forma, vários trabalhos defendem a utilização de mecanismos como métodos de XAI como formas de ultrapassar as limitações inerentes a redes neuronais profundas (Ribeiro et al. 2016). Por exemplo, Alamatsaz et al. (2024) aplicaram um modelo híbrido CNN-LSTM para a deteção de oito diferentes tipos de arritmias e obtiveram uma acurácia de 98.24%. No entanto, de modo a aumentar a confiança clínica e reduzir potenciais erros de diagnóstico, incorporaram o SHAP, que permitiu explicar a forma como o modelo fazia previsões.

2.2 Técnicas Explicáveis de Inteligência Artificial para interpretação de sinais de Eletrocardiograma (ECG)

A inteligência artificial explicável surge como forma de compreender, interpretar e inferir as previsões de um modelo de DL (Sreeja e Supriya 2023). Ou seja, XAI surge como um campo emergente que procura tornar as previsões das redes neuronais mais compreensíveis e confiáveis, aumentando a transparência e reduzindo a desconfiança nos mesmos (Singh

e Sharma 2022).

De acordo com Chaddad et al. (2023), os métodos XAI podem ser divididos em duas categorias principais:

- **Intrínsecos** – possuem uma estrutura simples, facilmente interpretável (por exemplo, modelos de regressão linear).
- ***Post hoc*** – fornecem informação interpretável obtida a partir de modelos já treinados, como redes neuronais.

Os métodos *post hoc* são, por isso, relevantes para modelos de DL, dada a sua complexidade. Estes métodos podem ser subdivididos de acordo com a sua explicabilidade em modelos específicos e modelos agnósticos (Chaddad et al. 2023). O âmbito de explicação pode ainda ser:

- **Local** – quando fornece interpretações para previsões individuais.
- **Global** – quando descreve o comportamento geral do modelo.

Assim, destacam-se dois métodos *post hoc* agnósticos:

Local Interpretable Model-Agnostic Explanations (LIME) – este método tem como objetivo fornecer uma interpretação do modelo original através da aproximação de um novo modelo mais simples, formado com base nas previsões de um modelo de DL (Ribeiro et al. 2016; Abdullah et al. 2023; Chaddad et al. 2023). Como é um método local, o LIME fornece explicações individualizadas, permitindo compreender quais características foram determinantes para uma dada previsão, mas não descreve o funcionamento global do modelo (Nguyen et al. 2021; Sreeja e Supriya 2023).

Desta forma, o processo baseia-se em introduzir perturbações controladas nos dados de entrada, gerar novas amostras ponderadas pela proximidade à instância analisada e treinar um modelo substituto interpretável. Este funcionamento é explicado pela seguinte equação (2.8):

$$\text{explain}(x) = \arg\min_{g \in G} \; L(f, g, \pi_x) + \Omega(g) \tag{2.8}$$

onde g representa o modelo interpretável, G o conjunto de modelos possíveis, L a função de fidelidade que avalia a aproximação ao modelo original e $\Omega(g)$ a penalização da complexidade para assegurar simplicidade e interpretabilidade (Ribeiro et al. 2016; Chaddad et al. 2023).

Shapley Additive Explanations (SHAP) – método baseado na teoria dos valores de Shapley, oriunda da Teoria dos Jogos, propõe explicar as previsões de um modelo para um determinado *input*, calculando a contribuição de cada característica para essa previsão (Lundberg e Lee 2017; Nguyen et al. 2021; Chaddad et al. 2023). Para tal, utiliza os valores SHAP, os quais correspondem aos valores da esperança condicional do modelo original, sendo, por isso, uma medida unificada da importância de cada característica (Lundberg e Lee 2017).

Como parte do enquadramento do SHAP, o *KernelSHAP* aproxima os valores de Shapley através de regressões lineares ponderadas. A sua principal contribuição é permitir calcular de forma eficiente a importância de cada *feature* para uma previsão específica (Lundberg e Lee 2017). Este modelo é definido pela equação (2.9):

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{2.9}$$

onde z' é um vetor binário que indica a presença ou ausência de *features*, e ϕ_i representa o valor de Shapley associado a cada *feature*. O peso de cada amostra é definido pelo *Shapley Kernel* (2.10):

$$\pi_{x'}(z') = \frac{M - 1}{\binom{M}{|z'|} \, |z'| \, (M - |z'|)} \tag{2.10}$$

em que M corresponde ao número total de *features* e $|z'|$ é o número de *features* presentes no subconjunto considerado. O termo $\binom{M}{|z'|}$ representa o coeficiente binomial, que indica o número de combinações possíveis de $|z'|$ elementos escolhidos entre M, sendo

definido pela equação 2.11:

$$\binom{M}{|z'|} = \frac{M!}{|z'|!\,(M - |z'|)!} \tag{2.11}$$

Este fator reflete o número de subconjuntos com igual dimensão e é utilizado para ponderar a contribuição de cada coligação no cálculo dos valores de Shapley.

De acordo com Lundberg e Lee (2017), comparativamente com o LIME, este método substitui escolhas heurísticas de *kernel* e a regularização por uma formulação teórica rigorosa, melhorando a eficiência amostral e garantindo explicações consistentes e alinhadas com a Teoria dos Jogos. Desta forma, apresenta-se na literatura como um dos principais métodos para explicar modelos de ML e DL.

Por conseguinte, a literatura recente demonstra a aplicabilidade destas abordagens em sinais ECG. Raza et al. (2022) desenvolveram um *framework*, baseado em CNN e XAI, para a classificação de arritmias cardíacas. Obtiveram uma precisão de 94.5% em sinais não filtrados e 98.9% em sinais filtrados da base de dados MIT-BIH Arrhythmia, evidenciando ganhos de transparência, redução de custos de comunicação e aumento na precisão da classificação.

De forma semelhante, Sreeja e Supriya (2023) integraram LIME e *Gradient-weighted Class Activation Mapping* (Grad-CAM) – como método XAI – a um modelo CNN treinado em imagens de ECG, alcançando um F1-score de 0.981 e demonstrando um aumento significativo no diagnóstico correto de DCV. Já Alamatsaz et al. (2024) aplicaram um modelo híbrido de CNN e LSTM para a deteção de arritmias cardíacas, complementado com SHAP e LIME, obtendo uma precisão de 98.24%.

Finalmente, Moningi et al. (2024) mostraram que o uso combinado de CNN-LSTM e métodos de XAI possibilita classificações de sinais com precisão acima de 98.25%. Tais resultados reforçam o aumento na interpretabilidade dos modelos de DL com explicações locais mais robustas e, por isso, tornam estes modelos mais confiáveis para uso clínico.

Em suma, a integração de XAI na análise de sinais ECG, para a classificação de DCV, representa um passo importante na direção de facilitar a leitura e interpretação de modelos de DL, transformando-os em ferramentas confiáveis e passíveis de uso na prática médica

(Chaddad et al. 2023; Sreeja e Supriya 2023).

2.3 Considerações éticas e regulamentares

A aplicação de técnicas de DL à classificação de sinais ECG tem demonstrado resultados significativos na deteção de DCV, constituindo uma oportunidade promissora para o avanço da medicina preditiva e personalizada (Chen et al. 2020; Chaddad et al. 2023; Zeng et al. 2024). No entanto, o sinal ECG é considerado um identificador biométrico único, uma vez que o seu traçado é característico de cada indivíduo (Ghazarian et al. 2022).

Por este motivo, os sinais ECG são considerados dados de saúde sensíveis e a sua recolha, tratamento, armazenamento e processamento requerem medidas robustas de salvaguarda da privacidade e da confidencialidade (Ghazarian et al. 2022). Entre estas medidas, destacam-se a anonimização e a pseudonimização dos dados de ECG, bem como a adoção de procedimentos éticos e regulamentares que assegurem a proteção do utente e a utilização responsável de informação médica (Ghazarian et al. 2022; Senthuran et al. 2025).

Na União Europeia, o Regulamento Geral sobre a Proteção de Dados (RGPD) – Regulamento (UE) 2016/679 – protege os indivíduos sempre que os seus dados sejam objeto de tratamento por entidades externas e estabelece os princípios fundamentais para o tratamento de dados pessoais e clínicos (artigos 1.º e 2.º). Entre estes princípios (artigo 5.º), salientam-se:

- **Pseudonimização** – medida técnica e organizativa destinada a reduzir os riscos de identificação dos doentes;
- **Licitude, lealdade e transparência** – exigência de tratamento lícito, justo e claro para com o titular;
- **Limitação de finalidades** – recolha de dados apenas para objetivos específicos, explícitos e legítimos, sem usos incompatíveis (exceto fins de investigação científica,

histórica ou estatística);

- **Minimização de dados** – utilização de informação adequada, pertinente e estritamente necessária;
- **Integridade e confidencialidade** – tratamento com medidas de segurança que evitem acessos não autorizados, perdas ou danos;
- **Responsabilidade** – obrigação do responsável pelo tratamento em garantir e demonstrar o cumprimento destes princípios.

A literatura sublinha a relevância destes aspetos, evidenciando os riscos associados ao uso de DL em sinais ECG. Ghazarian et al. (2022) mostraram que modelos convolucionais treinados em grandes bases de dados alcançaram 95.69% de precisão na identificação individual, demonstrando que mesmo os ECG anonimizados podem ser reidentificados, especialmente quando combinados com dados auxiliares, como a idade e o sexo. Já Senthuran et al. (2025) propuseram um *framework* para ambientes imersivos que procura integrar a proteção da privacidade com a integridade dos dados clínicos, concluindo que é viável implementar sistemas que simultaneamente preservem a confidencialidade do paciente e mantenham a utilidade clínica dos sinais ECG.

Capítulo 3

Metodologia

Neste capítulo descreve-se de forma detalhada todas as etapas do *pipeline* desenvolvido, representado esquematicamente na Figura 3.1. O modelo proposto consiste numa abordagem híbrida CNN-LSTM, aplicada à classificação de arritmias cardíacas a partir de sinais brutos de ECG. O método proposto inclui as fases de procura e recolha de bases de dados públicas, pré-processamento e catalogação dos sinais. Posteriormente, os vetores correspondentes às 12 derivações são introduzidos no modelo CNN-LSTM para discriminar entre três classes: batimento cardíaco normal, flutter auricular e outros ritmos. Para avaliar o desempenho da classificação, foram utilizados parâmetros de desempenho.

Por fim, de modo a conferir interpretabilidade ao modelo, foram aplicados métodos de XAI, nomeadamente SHAP e LIME.

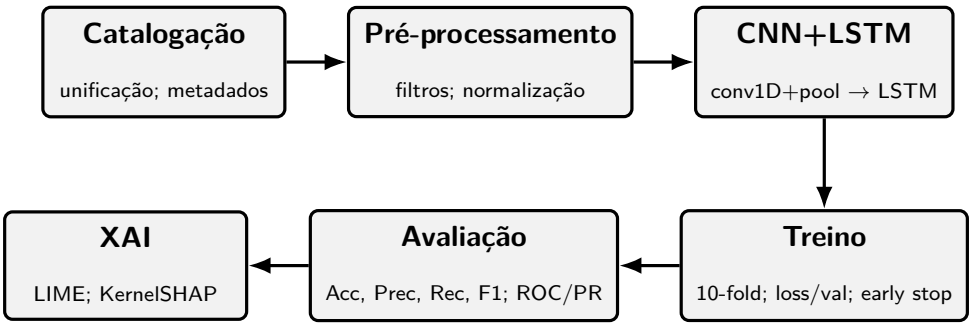


Figura 3.1: Esquematização do pipeline seguido no estudo

3.1 Base de dados

Na presente investigação, foram utilizadas três bases de dados públicas de sinais ECG de 12 derivações, disponibilizadas na plataforma *Physionet*, de modo a garantir diversidade de populações, condições clínicas e protocolos de aquisição.

Em todas as bases de dados, os dados são fornecidos no formado WFDB (*WaveForm DataBase format*), que se organiza em dois ficheiros (Perez Alday et al. 2020):

- **.hea (header)** – cabeçalho que contém, entre outros, metadados clínicos, incluindo sexo, idade, configuração das derivações e os códigos *SNOMED CT* (rótulos de classificação);
- **.mat (data)** – contém os dados binários em bruto correspondentes ao ECG.

Importa salientar que as bases de dados da *Physionet* são disponibilizadas já anonimizadas, em conformidade com princípios éticos e com o RGPD, assegurando o respeito pela privacidade dos utentes.

As bases de dados usadas, tanto para treino como para validação do modelo, foram as seguintes:

1. **A Large 12-Lead ECG Database for Arrhythmia Study**
   De acordo com (Zheng et al. 2020; Zheng et al. 2022), esta base de dados contém 45 152 registos de ECG de 12 derivações, com duração de 10 segundos e frequência de amostragem de 500 Hz. Inclui uma ampla variedade de patologias cardíacas, bem como registos de indivíduos saudáveis, sendo, por isso, adequada para estudos de classificação multiclasse.

2. **PTB-XL**
   Segundo (Wagner et al. 2020; Wagner et al. 2022), esta base de dados, proveniente do *Physikalisch Technische Bundesanstalt* (PTB), constitui uma extensão da PTB Diagnostic ECG Database. Contém 21 837 registos de ECG de 12 derivações, com 10 segundos de duração, registados a 500 Hz e 100 Hz, sendo 52% do sexo masculino

e 48% do sexo feminino. Em termos de utilização, literatura recente explora esta base de dados para estudos de *Deep Learning* e XAI na classificação de DCV (Anand et al. 2022).

3. **Georgia 12-Lead ECG Challenge Database**
   Esta base de dados com 10 344 observações, dos quais 54% correspondem a pacientes do sexo masculino e 46% do sexo feminino. Os sinais ECG têm 12 derivações, 10 segundos de duração e uma frequência de amostragem de 500 Hz. Tanto a PTB-XL como a Georgia 12-Lead ECG Challenge Database foram retiradas no âmbito do desafio *Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020* (Perez Alday et al. 2020; Perez Alday et al. 2022).

As bases de dados usadas nesta investigação incluem registos de pacientes com diagnósticos de uma única patologia e registos com múltiplas patologias associadas a um mesmo traçado de ECG. Nestes casos, uma única amostra de ECG pode apresentar um rótulo de classe única ou múltiplos rótulos em simultâneo (Anand et al. 2022).

Neste sentido, para garantir consistência na classificação, foram considerados apenas pacientes com diagnósticos de classe única, tendo sido filtrados e eliminados os registos com múltiplas classificações. Após este processo de seleção, a base de dados usada para treino, validação e teste ficou constituída por 31 864 observações.

Posteriormente, todos os sinais foram catalogados, convertidos para um formato unificado e guardados numa *struct* com os campos: `filename`, `signal`, `fs`, `acronym` e `target`. As variáveis idade (`age`) e sexo (`sex`) foram registadas, mas não incluídas no treino, para garantir que o classificador depende apenas da informação eletrofisiológica.

Assim, os sinais que compõem a base de dados apresentam a seguinte estrutura:

- Segmentos de 10 segundos a 500 Hz ($T = 5000$ amostras);
- Matriz $X \in \mathbb{R}^{12\times5000}$, correspondendo a 12 derivações (linhas) e 5000 amostras temporais (colunas);

- Três classes de saída: **Normal**, **AFL** e **Outros**. A classe Outros integra registos com ritmos cardíacos alterados, incluindo fibrilhação auricular, bloqueios de ramo e outras arritmias não AFL;
- Número total de registos: 31 864 (14 050 Normais, 1 494 AFL e 16 320 Outros)(Contagem dos 3 cenários na Figura 3.2);
- Metadados (idade, sexo e códigos clínicos) armazenados, mas não utilizados nesta fase do treino. A decisão de os excluir deve-se ao objetivo de avaliar exclusivamente a capacidade discriminativa dos sinais brutos de ECG, sem influência de fatores demográficos, e à inexistência de metadados em parte das bases de dados utilizadas, o que inviabilizaria a sua integração uniforme no processo de treino.

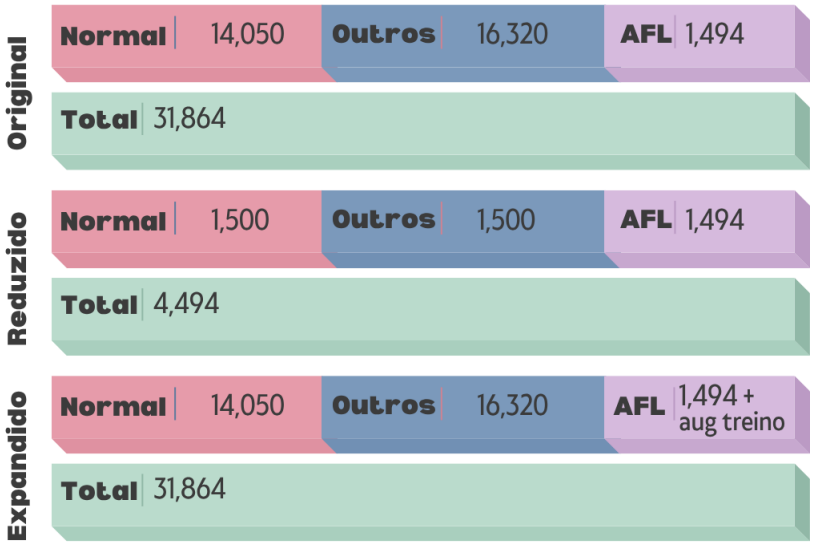


Figura 3.2: Total de Registos/Cenários

3.2 Pré-processamento dos sinais ECG

Os sinais ECG provenientes das bases de dados foram utilizados com 12 derivações padrão (I, II, III, aVR, aVL, aVF, V1 a V6), e representados como matrizes bidimensionais $X \in \mathbb{R}^{12 \times T}$, correspondentes a segmentos de 10 segundos a 500 Hz.

Com o objetivo de remover interferências, reduzir o ruído e garantir consistência na escala dos dados, foi aplicado um processo de pré-processamento uniforme a cada uma das 12 derivações. O pré-processamento seguiu as seguintes etapas:

1. **Filtragem digital multicanal**

   Cada derivação foi filtrada através da combinação de três filtros digitais:

   - Um filtro passa-baixo de Butterworth, com frequência de corte de 50 Hz, calculada automaticamente com base em critérios de atenuação de 3 dB na banda de passagem e 60 dB na banda de rejeição. Esta filtragem elimina ruídos de alta frequência, como os provenientes de contrações musculares ou interferências técnicas (Zheng et al. 2020);
   - Dois filtros *Notch* (rejeita-banda), centrados em 50 Hz e 100 Hz, com fator de qualidade $Q = 30$, utilizados para suprimir a interferência da rede elétrica (corrente alternada) e do seu primeiro harmónico, respetivamente.

   A aplicação combinada destes três filtros foi realizada por convolução dos respetivos coeficientes, originando um sistema de filtragem equivalente que foi aplicado a cada canal, de forma a garantir uma implementação eficiente e uniforme.

2. **Remoção do desvio de base (*baseline drift*)**

   Após a filtragem, foi removido o desvio de base de cada canal através da subtração da média do sinal. Este procedimento é responsável pela eliminação de componentes de baixa frequência capazes de provocar alterações da linha de base, como artefactos causados por movimento, respiração ou o posicionamento dos elétrodos. Desta forma, as ondas do ECG são fielmente representadas (Ganeshkumar et al. 2023).

3. **Normalização da amplitude**

   Finalmente, cada derivação foi normalizada para o intervalo $[-1, 1]$, com recurso à divisão do sinal filtrado e centrado pelo seu valor máximo absoluto. Este procedimento assegura que todos os registos, independentemente da amplitude original, apresentem uma escala relativa comum, o que facilita a generalização dos modelos de DL e evita o enviesamento do treino (Ping et al. 2020).

3.3 *Data Augmentation*

De forma a mitigar o desbalanceamento entre classes, em particular a reduzida representação da classe flutter auricular (AFL), foi aplicado um conjunto de técnicas de *data augmentation* aos registos de treino do cenário Expandido. Estas transformações foram implementadas durante o processo de treino e apenas à classe AFL, mantendo os conjuntos de validação e teste inalterados. O objetivo foi aumentar a variabilidade dos sinais e melhorar a capacidade de generalização do modelo CNN-LSTM, preservando simultaneamente a morfologia essencial do traçado do sinal ECG (Hatamian et al. 2020; Raghu et al. 2022).

As transformações aplicadas foram as seguintes:

- **Adição de ruído branco:** introdução de ruído gaussiano com média zero e desvio padrão de 0.003 sobre o sinal normalizado. Esta perturbação simula interferências fisiológicas ou técnicas (como ruído muscular ou de elétrodo), sem comprometer a estrutura da onda;

- **Variação de ganho:** multiplicação do vetor de amostras por um fator aleatório normalmente distribuído, $\alpha \sim \mathcal{N}(1, 0{,}01^2)$, aplicado individualmente a cada derivação de modo a reproduzir variações na amplitude de registo entre registos;

- **Deslocamento temporal:** rotação circular do vetor de amostras até ± 20 amostras (± 0.04 s a 500 Hz), mantendo a continuidade do sinal e simulando pequenas diferenças de sincronização entre batimentos cardíacos;

Estas operações foram aplicadas estocasticamente a cada iteração de treino, com probabilidades de 0.6, 0.4 e 0.3 para ruído, variação de ganho e deslocamento, respetivamente. Assim, o número de amostras permaneceu inalterado, mas a cada época eram geradas novas variações do sinal, promovendo maior robustez do modelo e reduzindo o viés associado ao desbalanceamento de classes.

Com esta estratégia, pretendeu-se reduzir o viés resultante do desbalanceamento de classes, aumentar a robustez da rede neuronal e permitir que o modelo aprenda representações mais gerais dos padrões característicos do flutter auricular.

3.4 Arquitetura de Deep Learning

Nesta investigação foi desenvolvida uma abordagem híbrida CNN-LSTM para a classificação de DCV em sinais ECG de 12 derivações. De acordo com a literatura, esta arquitetura tem demonstrado consistentemente resultados significativos na identificação correta de patologias cardíacas (Yao et al. 2020; Zeng et al. 2023), uma vez que combina a capacidade de extrair representações morfológicas com a modelação de dependências temporais nos sinais.

De acordo com Yıldırım et al. (2018), foi adotada uma estratégia de aprendizagem profunda *end-to-end*, na qual os sinais ECG, após pré-processamento, foram utilizados diretamente como entrada para as redes neurais (CNN e LSTM), sem necessidade de extração manual de *features*. Assim, a arquitetura proposta foi desenvolvida com base numa combinação de camadas convolucionais unidimensionais (1D CNN) e camadas recorrentes LSTM, seguidas por uma cabeça densa (*fully connected*), responsável pela decisão de classe. A rede foi implementada em `MATLAB`, com recurso à `Deep Learning Toolbox` para a modelação, treino e avaliação, recebendo como entrada sequências multivariadas de dimensão $\mathbb{R}^{12\times T}$, correspondendo às 12 derivações padrão, sendo T o número de amostras no tempo.

A estrutura da arquitetura está representada na Figura 3.3 e está dividida em três blocos principais:

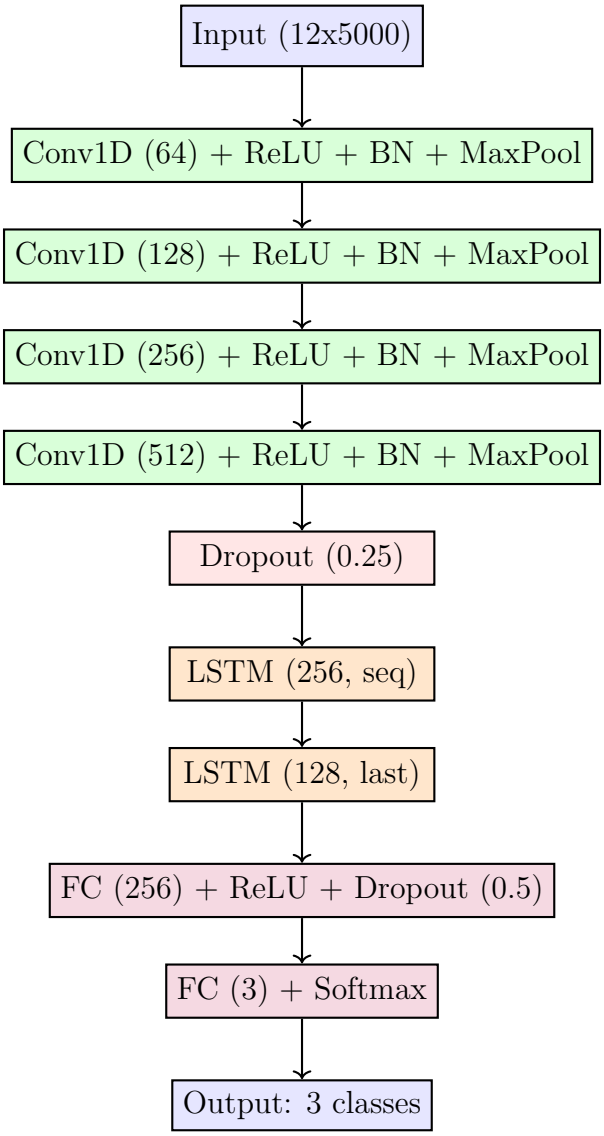


Figura 3.3: Arquitetura da rede CNN+LSTM profunda utilizada para classificação dos sinais ECG em 12 derivações.

1. **Bloco Convolucional (1D CNN)**

   Foi utilizada uma CNN unidimensional, adequada à natureza temporal dos sinais ECG, em contraste com as CNN 2D, que operam em duas dimensões e são tradicionalmente aplicadas a imagens (Yıldırım et al. 2018; Zeng et al. 2023).

   Portanto, seguindo a estruturação de Zeng et al. (2023), mas adaptando para sinais com 12 derivações, este bloco tem como objetivo extrair características espaciais

locais ao longo da dimensão temporal de cada canal ECG. Inclui quatro camadas convolucionais sequenciais, intercaladas com normalização, função de ativação e *pooling*. A primeira camada Conv1d(5, 64) aplica 64 filtros de ativação com *kernel* de tamanho 5, de forma a capturar padrões morfológicos curtos; a segunda camada Conv1d(5, 128) aumenta a profundidade da representação com 128 filtros e *kernel* de tamanho 5; a terceira camada Conv1d(3, 256) introduz 256 filtros com *kernel* menor (3 amostras), permitindo refinar detalhes locais; e, por último, a camada Conv1d(3, 512), que utiliza 512 filtros e *kernel* de tamanho 3, obtendo uma representação abstrata e profunda do sinal antes das camadas de LSTM.

Cada convolução é seguida por *Batch Normalization* (Ioffe e Szegedy 2015; Yao et al. 2020) para estabilização do treino, ReLU (Nair e Hinton 2010; Yao et al. 2020; Sreeja e Supriya 2023), que previne o problema do gradiente em redes neurais profundas, e *Max Pooling* com fator 2, que reduz a dimensionalidade temporal da amostra, evitando *overfitting* e permitindo um treino mais rápido.

2. **Bloco Recorrente (LSTM)**

   O objetivo deste bloco é capturar dependências temporais de longo prazo e padrões sequenciais ao longo do tempo, após a extração de características convolucionais.

   Este bloco é composto por duas camadas LSTM em cascata, com diferentes modos de saída: a primeira camada LSTM com 256 unidades, que processa a sequência de vetores de ativação resultantes da CNN, de forma recorrente, e retorna um vetor de saída para cada instante (*timestep*) da sequência. E a segunda camada LSTM com 128 unidades, na qual apenas o vetor final de estado oculto da sequência é utilizado como saída. Neste sentido, este vetor final representa uma codificação compacta e contextualizada da sequência completa.

3. **Bloco Denso (Classificador)**

   Por fim, o bloco de classificação totalmente conectado (*fully connected*) transforma o vetor final da LSTM num vetor de probabilidades associado às diferentes classes de classificação. Sendo, por isso, a camada responsável pela decisão de classe.

A primeira camada deste bloco é uma camada densa com 256 neurónios, seguida por uma função de ativação ReLU e uma camada de Dropout (0.5). Esta camada atua como camada oculta e tem como objetivo introduzir uma não-linearidade e evitar sobreajuste (*overfitting*).

Em seguida, é utilizada uma camada densa final (*fullyConnected*) com um número de unidades igual ao número de classes, no caso três classes, e produz um vetor de *logits*, que, por sua vez, é transformado em probabilidades normalizadas pela função *Softmax*.

Finalmente, a classificação é dada com recurso a uma função de classificação, que compara as probabilidades previstas com os rótulos reais. Adicionalmente, para mitigar o efeito do desbalanceamento entre classes, foram adicionados pesos de classe ajustados em função da frequência de ocorrência, isto é, daqueles que apresentam menos exemplos.

3.5 Estratégia de Treino, Validação e Métricas de Avaliação

3.5.1 Treino, Validação e Teste

Após a definição da arquitetura da rede neuronal, foram considerados três cenários para treino do modelo DL: base de dados original, base de dados reduzida (limitando as classes Normal e Outros a 1500 sinais cada, mantendo a AFL completa) e base de dados expandida, obtida através de técnicas de *data augmentation* (Hatamian et al. 2020; Raghu et al. 2022) aplicadas exclusivamente à classe AFL, nas quais foram usadas, apenas nos dados de treino, adição de ruído branco, variação de ganho e deslocamento temporal.

Para avaliar o desempenho de forma robusta, foi utilizada a técnica de validação cruzada a 10 *folds* (Faust et al. 2018; Zeng et al. 2024). Em cada iteração, o conjunto de dados foi dividido em 10 subconjuntos de igual dimensão, como mostra a figura 3.4: 9 foram utilizados para treino e validação e 1 para teste. O processo foi repetido 10 vezes, alternando os subconjuntos, e os resultados médios foram calculados sobre todos os *folds*. Este método assegura que todas as amostras são utilizadas tanto para treino/validação como para teste, evitando dependência de uma única divisão dos dados e reduzindo o risco de sobreajuste.

Durante o treino de cada *fold*, a função de custo de validação (*validation loss*) foi monitorizada a cada época para selecionar o modelo com melhor desempenho local (Sreeja e Supriya 2023).

O treino foi executado durante um máximo de 100 épocas, com recurso a *early stopping* com paciência de 10 épocas, ou seja, o treino era interrompido automaticamente se não se verificasse melhoria na *validation loss* durante 10 épocas consecutivas, mitigando o risco de *overfitting* (Sreeja e Supriya 2023). O *optimizer* usado foi o Adam (Kingma e Ba 2014), com parâmetros *default* e um *learning rate* de 0.0001 (para estabilizar a convergência e minimizar o impacto de classes desbalanceadas, evitando que o modelo privilegie padrões de classes com maior frequência), de modo a reduzir os efeitos causados

pelo desbalanceamento dos dados (Yao et al. 2020).

O treino/validação e teste do modelo foram realizados numa máquina virtual disponibilizada pelo Instituto Politécnico de Bragança, com processador de 16 núcleos a 3.0 GHz, 64 Gb de memória RAM, armazenamento SSD de 500 GB e uma GPU dedicada com 16 GB de memória dedicada com suporte a CUDA.

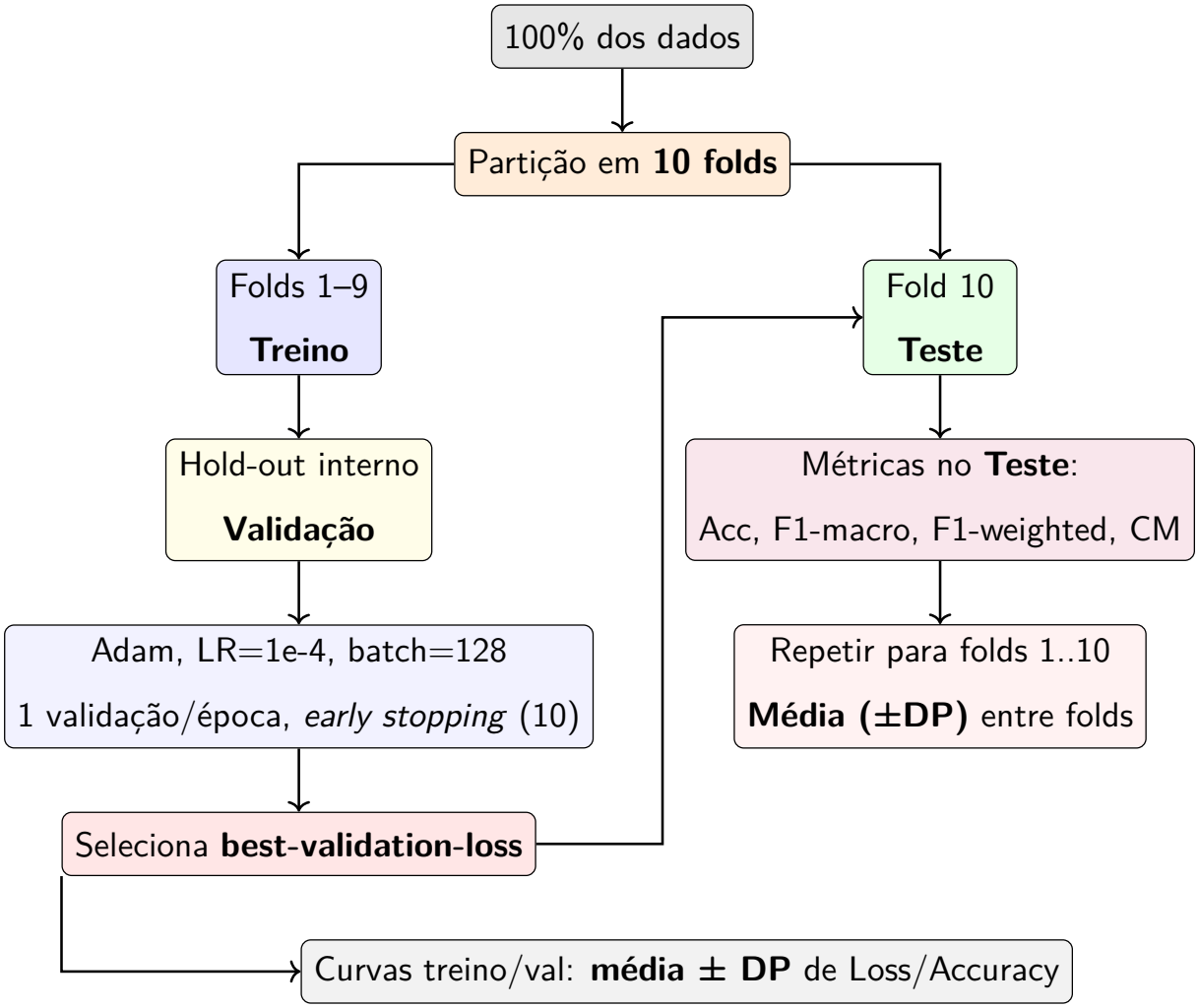


Figura 3.4: Validação cruzada 10-fold com um fold de teste por iteração e validação interna 80/20 no treino.

3.5.2 Métricas de Avaliação

A avaliação do desempenho dos modelos foi conduzida com base em métricas de classificação multiclasse, amplamente reportadas na literatura na análise de sinais biomédicos (Yao et al. 2020; Raza et al. 2022; Ganeshkumar et al. 2023).

Acurácia (*Accuracy*): mede a proporção de previsões corretas em relação ao total de observações (Anand et al. 2022; Singh e Sharma 2022; Raza et al. 2022; Ganeshkumar et al. 2023).

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Onde, TP (*True Positive*) é o número de amostras corretamente classificadas como positivas, TN (*True Negative*) é o número de amostras classificadas corretamente como negativas, FP (*False Positive*) é o número de amostras classificadas incorretamente como positivas e FN (*False Negative*) é o número de amostras positivas classificadas como negativas.

Precisão (*Precision*): mede a proporção de exemplos classificados como positivos que são de facto corretos (Anand et al. 2022; Singh e Sharma 2022; Raza et al. 2022; Ganeshkumar et al. 2023).

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{3.2}$$

Sensibilidade (*Recall*): mede a capacidade do modelo em identificar corretamente os exemplos positivos (Anand et al. 2022; Singh e Sharma 2022; Raza et al. 2022; Ganeshkumar et al. 2023).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.3}$$

Especificidade (*Specificity*): mede a capacidade do modelo em identificar corretamente exemplos negativos.

$$\text{Especificidade} = \frac{TN}{TN + FP} \tag{3.4}$$

F1-Score: corresponde à média harmónica entre a Precisão e a Sensibilidade, equilibrando as duas métricas (Anand et al. 2022; Singh e Sharma 2022; Raza et al. 2022;

Ganeshkumar et al. 2023).

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \tag{3.5}$$

AUC (*Area Under Curve*): métrica associada à curva ROC (*Receiver Operating Characteristic*), a qual representa a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) para diferentes limiares de decisão. O AUC corresponde à área sob a curva ROC e varia entre 0 e 1, refletindo a capacidade discriminativa global do modelo (Anand et al. 2022).

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{TN + FP} \tag{3.6}$$

Average Precision (AP): métrica associada à curva *Precision–Recall* (PR) e corresponde à área sob a curva (AUC–PR). Resume a relação entre precisão e *recall* para todos os limiares de decisão, refletindo a capacidade do modelo em manter uma elevada precisão mesmo quando o *recall* aumenta. É particularmente útil em problemas com classes desbalanceadas, como a deteção de arritmias raras, sendo definida pela soma das precisões ponderadas pelas variações sucessivas de *recall*:

$$AP = \sum_{n=1}^{N} (R_n - R_{n-1}) \times P_n \tag{3.7}$$

onde P_n e R_n representam, respetivamente, a precisão e o *recall* no n-ésimo ponto da curva PR.

Dado que o conjunto de dados utilizado apresenta desbalanceamento entre classes, recorreu-se a técnicas de média para obter métricas mais representativas (Anand et al. 2022). Assim foram usadas:

Macro-averaging: calcula cada métrica por classe e, em seguida, encontra a média aritmética simples entre classes. Todas as classes têm o mesmo peso.

$$M_{\text{macro}} = \frac{1}{C} \sum_{i=1}^{C} \text{Métrica}_i \tag{3.7}$$

Weighted-averaging: calcula a métrica por classe e aplica a média ponderada pelo número de amostras de cada classe. Assim, o peso é proporcional à sua representatividade no conjunto total.

$$M_{\text{weighted}} = \frac{1}{N} \sum_{i=1}^{C} n_i \cdot \text{Métrica}_i \tag{3.8}$$

Onde C é o número total de classes, n_i é o número de amostras da classe i e N o número total de amostras.

Capítulo 4

Resultados

Neste capítulo são apresentados e discutidos os testes realizados, com o objetivo de verificar se o projeto cumpre os objetivos propostos e se os classificadores desenvolvidos conseguem distinguir eficazmente diferentes classes de ECG.

Foram implementados três cenários distintos:

1. **Dataset Original**: treino com todos os sinais disponíveis, mantendo o desbalanceamento natural entre classes;
2. **Dataset Reduzido**: classes 0 (Normal) e 2 (Outro) reduzidas para 1500 sinais cada, mantendo a classe 1 (AFL) com os registos originais;
3. **Dataset Expandido**: treino com aumento de dados (*data augmentation*) na classe 1, de modo a reforçar a representação de *Atrial Flutter*.

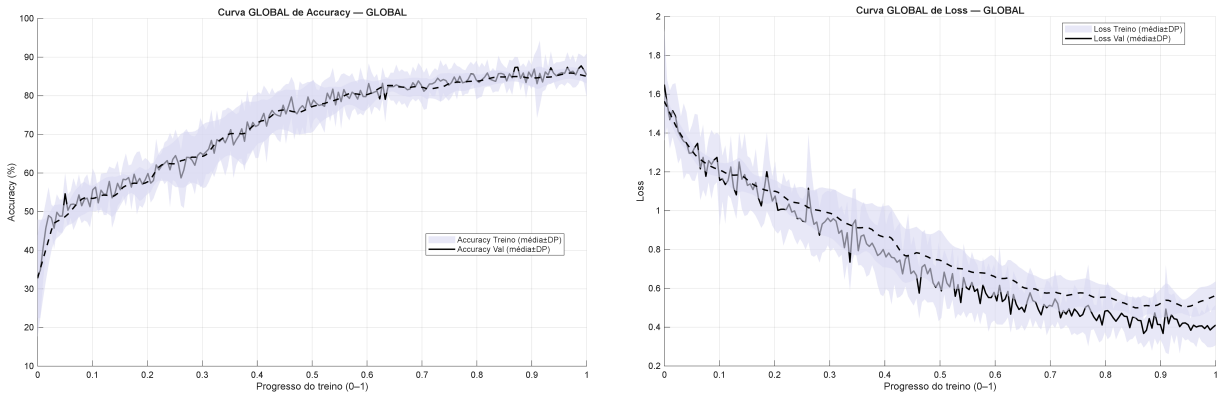
4.1 Treino e Validação

As Figuras 4.1a e 4.1b apresentam a evolução global da acurácia e a função de custo *loss* ao longo do processo de treino, agregada sobre os 10 *folds* de validação cruzada. As linhas representam a média e as áreas sombreadas indicam o desvio padrão entre folds.

Neste sentido, é possível visualizar que a acurácia nos conjuntos de treino e validação evoluem de forma paralela e convergem de forma estável. Este comportamento evidencia a

capacidade de generalização do modelo, dado que não se verificam diferenças substanciais entre o desempenho em treino e em validação, nem sinais de overfitting.

Por sua vez, a curva de custo revela uma diminuição consistente tanto no treino como na validação, estabilizando em valores baixos e sem tendência de divergência. Este resultado reforça a robustez do modelo e confirma que a configuração adotada permite uma aprendizagem eficaz e consistente.



(a) Curva global de Accuracy (10-fold) (b) Curva global de Loss (10-fold)

Figura 4.1: Curvas globais de desempenho médio ao longo do treino num processo de validação cruzada 10-folds. As áreas sombreadas representam o desvio padrão entre folds.

4.2 Resultados dos Cenários

A Tabela 4.1 mostra os resultados médios das métricas globais calculados a partir da validação cruzada a 10 folds para cada um dos cenários considerados. Verifica-se que os cenários treinados com os dados originais e com os dados expandidos alcançam desempenhos semelhantes e, em geral superiores aos obtidos com os dados reduzidos. Estes resultados podem ser explicados, em certa medida, pela menor quantidade de sinais disponíveis na base de dados Reduzida, o que compromete a capacidade de generalização do modelo.

Entre as métricas, destaca-se a precisão ($\text{Prec}_{\text{macro}}$), em que os cenários Original e Expandido atingem valores próximos de 0.81, enquanto o cenário Reduzido apresenta um valor inferior (0.755). Da mesma forma, o F1-Score é de aproximadamente 0.84 para os

cenários Original e Expandido, contra 0.75 no cenário Reduzido. Apesar dos resultados promissores, os valores reportados ainda se situam abaixo dos estudos prévios, como os de (Chen et al. 2020; Madan et al. 2022), onde com abordagens CNN-LSTM obtiveram precisões superiores a 95% em bases de dados específicas.

Dataset	Acc	$\text{Prec}_{\text{macro}}$	$\text{Rec}_{\text{macro}}$	$\text{Spec}_{\text{macro}}$	F1_{macro}	$\text{AUC}_{\text{macro}}$	AP_{macro}
Original	0.891	0.805	0.908	0.940	0.840	0.972	0.882
Expandido	0.891	0.806	0.912	0.940	0.843	0.972	0.881
Reduzido	0.754	0.755	0.754	0.877	0.752	0.892	0.801

Tabela 4.1: Comparação global dos modelos (médias 10-fold)

A Tabela 4.2 mostra a acurácia obtida em cada fold da validação cruzada 10-fold para os três cenários. Observa-se uma elevada consistência entre folds nos cenários Original e Expandido, cujos valores se mantém estáveis, entre aproximadamente 87% e 91%. Já o cenário Reduzido sofre maior variabilidade e perda de desempenho, com acurácias a oscilar entre 69.7% (fold 8) e 79.5% (fold 1, melhor desempenho).

Neste sentido vemos que, no cenário Original destaca-se a fold 9, onde 91.3% das previsões estão corretas, valor alinhado com o comportamento global. Enquanto na melhor fold do cenário Reduzido, apenas 79.5% das previsões estão corretas.

Estes resultados estão em consonância com a literatura, Yıldırım et al. (2018) demonstraram que uma abordagem end-to-end baseada numa 1D-CNN obteve acurácias superiores a 90%. Zeng et al. (2023) reportaram também valores superiores a 90% utilizando uma arquitetura CNN-LSTM, mas com extração de features e a apenas uma derivação de ECG. Tais evidências reforçam a robustez dos cenários Original e Expandido e a limitação do cenário Reduzido em termos de capacidade de generalização.

Dataset	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10
Original	0.887	0.884	0.881	0.890	0.892	0.886	0.875	0.900	0.913	0.901
Expandido	0.873	0.905	0.898	0.881	0.880	0.883	0.880	0.909	0.900	0.903
Reduzido	0.795	0.748	0.755	0.739	0.782	0.759	0.724	0.697	0.780	0.764

Tabela 4.2: Accuracy por fold (10-fold) dos três modelos

As Tabelas 4.3, 4.4 e 4.5 detalham o desempenho obtido por classe em cada um dos cenários avaliados.

Dataset	Prec	Rec	Spec	F1	AUC	AP
Original	0.578	0.950	0.965	0.716	0.987	0.722
Expandido	0.584	0.961	0.966	0.725	0.988	0.719
Reduzido	0.833	0.879	0.910	0.853	0.953	0.893

Tabela 4.3: Comparação por classe **Afl** (médias 10-fold).

Dataset	Prec	Rec	Spec	F1	AUC	AP
Original	0.925	0.913	0.931	0.919	0.973	0.970
Expandido	0.919	0.918	0.924	0.918	0.974	0.970
Reduzido	0.738	0.714	0.873	0.724	0.887	0.781

Tabela 4.4: Comparação por classe **Normal** (médias 10-fold).

Dataset	Prec	Rec	Spec	F1	AUC	AP
Original	0.911	0.862	0.925	0.885	0.955	0.953
Expandido	0.916	0.857	0.930	0.885	0.956	0.954
Reduzido	0.695	0.670	0.848	0.679	0.836	0.731

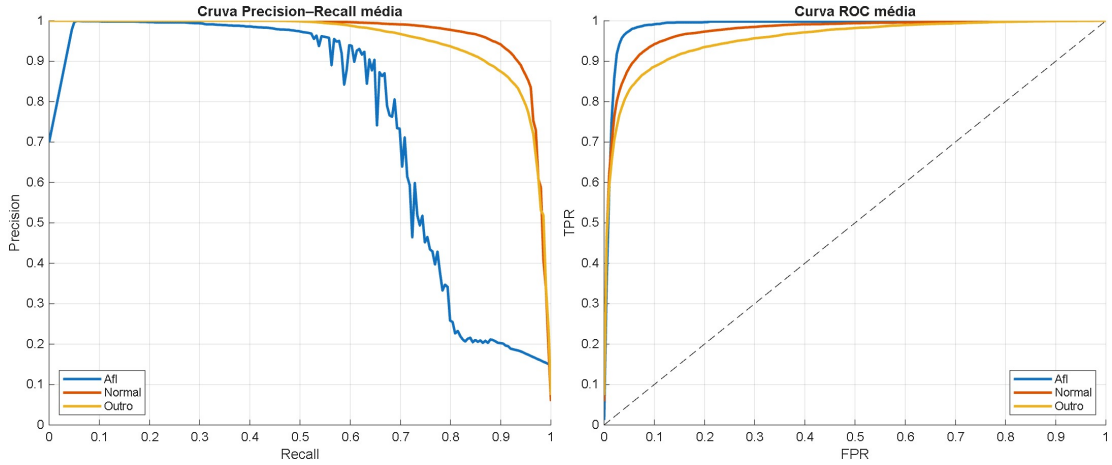
Tabela 4.5: Comparação por classe **Outros** (médias 10-fold).

Para a classe AFL, observa-se que os cenários Original e Expandido alcançam valores de Recall muito elevados (0.950 e 0.961, respetivamente), demonstrando uma elevada sensibilidade, isto é, conseguem detetar a grande maioria dos episódios reais de AFL. No entanto, este desempenho é acompanhado por valores de precisão relativamente baixos (0.578 e 0.584), revelando uma maior incidência de falsos positivos.

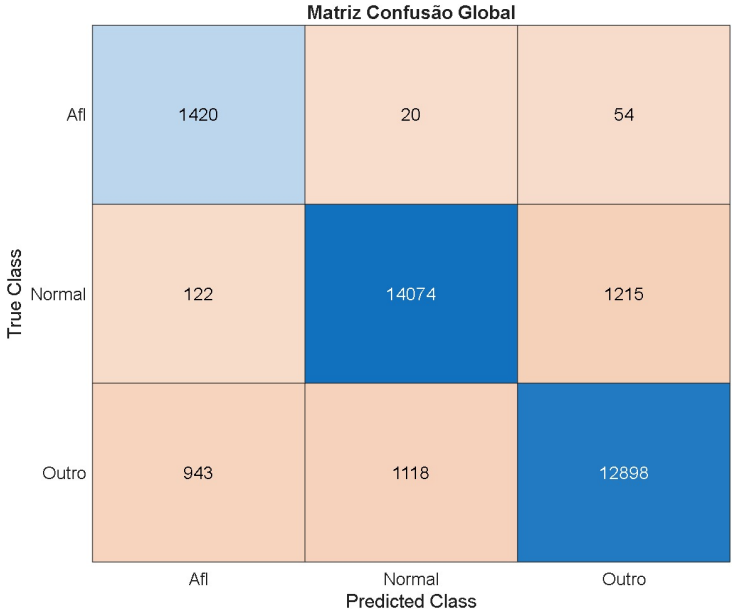
Por outro lado, o cenário Reduzido apresenta um comportamento oposto, atinge uma precisão superior (0.833), indicando que a maioria dos sinais classificados como AFL são efetivamente positivos. Contudo, o Recall (0.879) é inferior ao dos cenários Original e Expandido, o que significa que falha na identificação de alguns episódios reais de AFL.

Em termos globais, refletidos pelo F1-Score, o cenário Reduzido (0.853) obtém melhor equilíbrio entre Precisão e Recall, superando os valores dos cenários Original (0.716) e Expandido (0.725). Estes resultados evidenciam que a AFL constitui a classe mais difícil de classificar corretamente, devido à sua morfologia variável. No entanto, a análise evidencia que cada cenário favorece um compromisso diferente entre precisão e Recall: os cenários Original e Expandido são mais sensíveis à identificação da AFL, enquanto o cenário Reduzido é mais específico, mas menos abrangente.

As Figuras 4.2, 4.3, 4.4 mostram os resultados das matrizes confusão e das curvas PR e ROC de cada cenário. Desta forma é possível observar que os resultados apresentados em cima são consistentes com os obtidos nas matrizes confusão. No cenário Reduzido, para a classe AFL (Figura 4.3 (c)), de um total de 1584 sinais classificados como AFL, 1314 foram classificados corretamente (True Positive), correspondendo a uma precisão próxima de 83%. Por sua vez, as Curvas PR (Precisão-Recall) confirmam a mesma tendência, isto é, o cenário Reduzido atinge maior precisão, mas a um custo de menor Recall, enquanto os cenários Original e Expandido conseguem captar uma maior proporção de episódios reais de AFL.

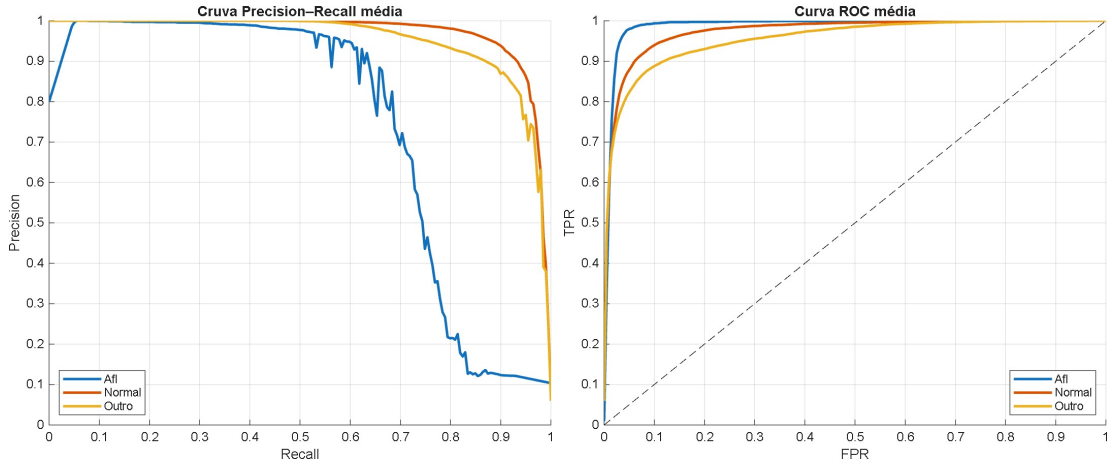


(a) Curva Precision-Recall (b) Curva ROC

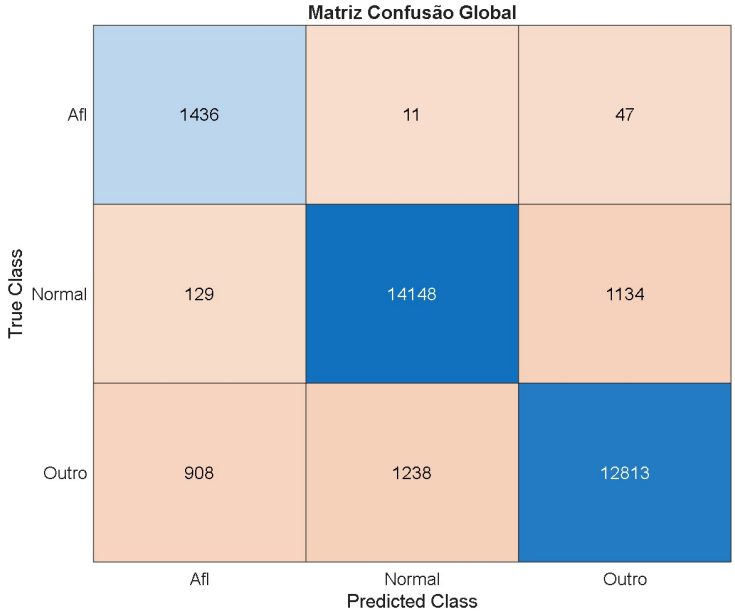


(c) Matriz Confusão

Figura 4.2: Resultados das Curvas e Matriz confusão do dataset original

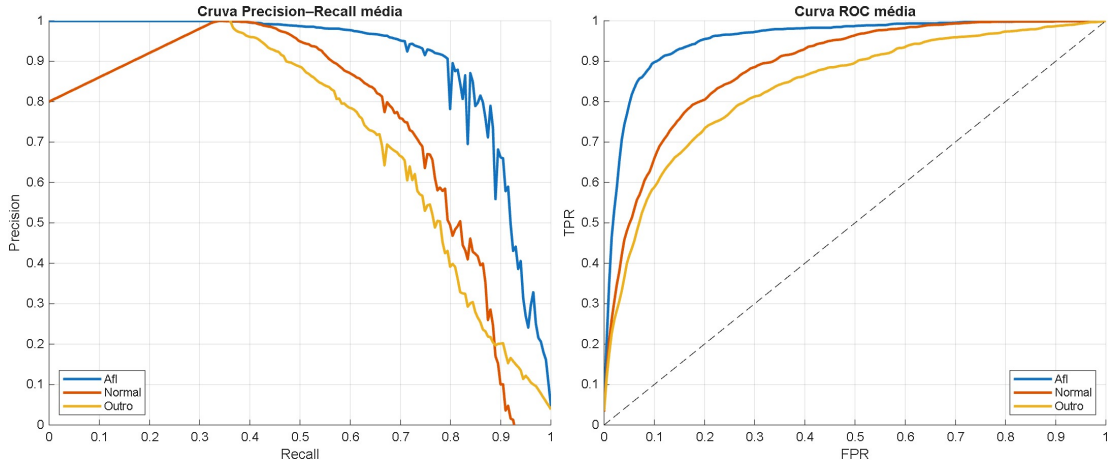


(a) Curva Precision-Recall (b) Curva ROC

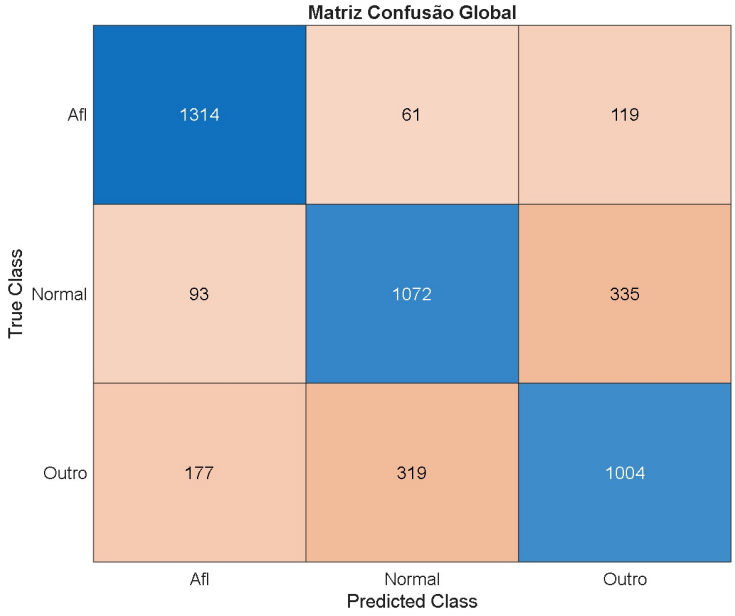


(c) Matriz Confusão

Figura 4.3: Resultados das Curvas e Matriz confusão do dataset expandido



(a) Curva Precision-Recall (b) Curva ROC



(c) Matriz Confusão

Figura 4.4: Resultados das Curvas e Matriz confusão do dataset reduzido

4.3 Explainable AI - SHAP e LIME

Apesar dos modelos de DL apresentarem resultados promissores na classificação de doenças cardiovasculares (DCV), continuam frequentemente a ser designados como "*black boxes*", devido à falta de interpretabilidade no modo como processam os sinais e chegam às

previsões (Anand et al. 2022). Para colmatar esta limitação, têm sido propostos métodos de XAI (Ribeiro et al. 2016; Singh e Sharma 2022; Alamatsaz et al. 2024), entre os quais se destacam o SHAP (Lundberg e Lee 2017) e o LIME (Ribeiro et al. 2016).

Nesta investigação, foi utilizado o *KernelSHAP* para avaliar a contribuição relativa de cada derivação do ECG no processo de classificação do modelo. Cada derivação foi considerada como uma *feature* independente, permitindo quantificar o impacto individual de cada uma na probabilidade predita pelo modelo. Assim, para cada exemplo, foi obtido um vetor de 12 valores SHAP (um por derivação), que expressa o contributo positivo ou negativo de cada derivação para a decisão final.

Tal como ilustrado no gráfico de violino da Figura 4.5, esta análise revelou que as derivações II, III, aVF e V1 são as que mais influenciam a classificação da classe *Atrial Flutter* (AFL), em concordância com a literatura (Cosio 2017). A derivação II evidencia a morfologia típica em "dente de serra" associada ao *flutter*, enquanto a V1 é sensível às variações na condução atrioventricular. As restantes derivações apresentam valores SHAP próximos de zero, sugerindo menor relevância para o modelo, o que demonstra coerência fisiológica na forma como a rede CNN–LSTM interpreta os sinais.

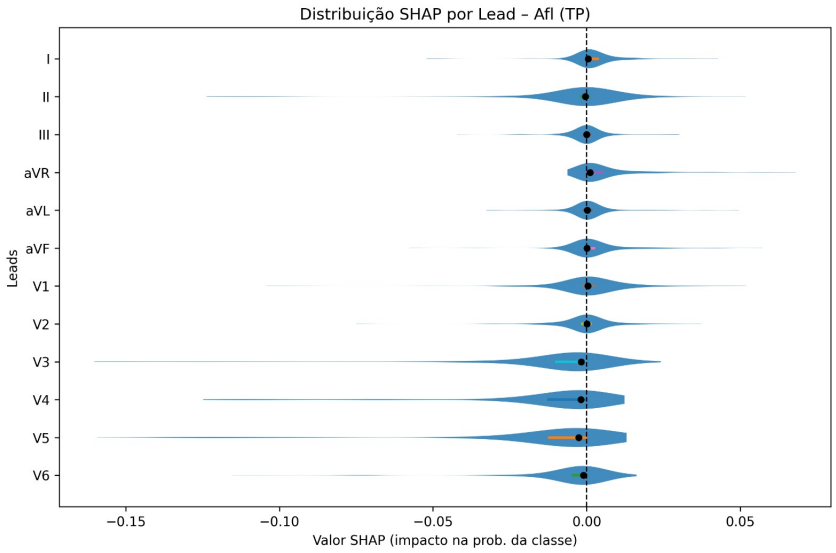


Figura 4.5: Distribuição dos valores SHAP por derivação para a classe flutter auricular (Afl–TP).

Adicionalmente, foi aplicado o método LIME para uma interpretação local do modelo. Neste caso, o registo de ECG de 12 derivações foi segmentado em janelas de 1 segundo, sobre as quais foi ajustada uma regressão linear de forma a aproximar localmente o comportamento do classificador. Os resultados obtidos, apresentados na Figura 4.6, são representados sob a forma de mapas de calor (*heatmaps*) que evidenciam os segmentos temporais com maior contributo para a decisão do modelo.

De forma global, o LIME apontou as derivações II, aVF e V1 como as mais relevantes para a deteção de flutter auricular, reforçando as conclusões obtidas pelo SHAP. No entanto, observa-se que algumas zonas de cor intensa ocorrem em segmentos do sinal que, do ponto de vista clínico, não correspondem a morfologias características do flutter. Este comportamento resulta da natureza local e perturbativa do LIME, que estima a importância dos segmentos através de variações estatísticas, podendo assim sobrevalorizar regiões que coincidem temporalmente com padrões relevantes, mas que não apresentam significado fisiológico direto. Ainda assim, a concentração de pesos elevados nas derivações II, aVF e V1 confirma que o modelo baseia a sua decisão em regiões fisiologicamente plausíveis.

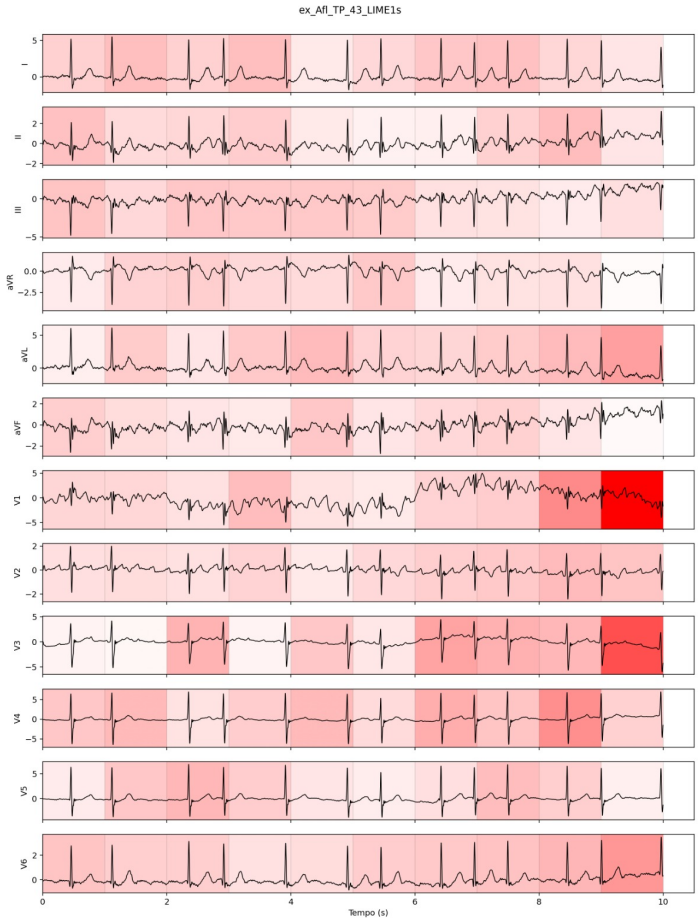


Figura 4.6: Mapa de calor LIME para um exemplo verdadeiro positivo da classe *Atrial Flutter* (Afl–TP).

A análise conjunta dos métodos SHAP e LIME evidencia a sua complementaridade. O SHAP fornece uma visão global da importância média de cada derivação em todo o conjunto de dados, enquanto o LIME oferece uma perspetiva local, permitindo identificar os segmentos temporais mais determinantes em exemplos individuais. Apesar de pequenas discrepâncias locais, os dois métodos convergem na identificação das mesmas derivações-chave, reforçando a coerência fisiológica da explicabilidade do modelo e a sua potencial aplicabilidade em contextos clínicos.

Capítulo 5

Discussão Geral e Conclusões

5.1 Discussão

Os resultados obtidos neste estudo evidenciam o potencial das arquiteturas híbridas CNN–LSTM na classificação automática de arritmias cardíacas a partir de sinais brutos de ECG. O modelo proposto atingiu uma acurácia média próxima de 89% e um F1-*score* macro de cerca de 0.84 nos cenários Original e Expandido, valores que se encontram de acordo com os reportados por Yıldırım et al. (2018), Chen et al. (2020) e Madan et al. (2022) em trabalhos semelhantes. Estes resultados confirmam que a combinação entre redes CNN (responsáveis pela extração espacial de padrões), e redes LSTM, que capturam dependências temporais, é eficaz na representação das dinâmicas eletrofisiológicas dos sinais cardíacos.

A análise comparativa entre os três cenários de treino permitiu avaliar o impacto do desbalanceamento de classes e das técnicas de *data augmentation*. O cenário Reduzido apresentou um desempenho global inferior (F1-*score* = 0.75), mas melhor desempenho na classe flutter auricular (Precisão = 0.83), demonstrando que a redução de segmentos das classes maioritárias aumenta a sensibilidade para classes minoritárias, ainda que à custa de uma menor estabilidade global. Por outro lado, o cenário Expandido – obtido com técnicas de *data augmentation* aplicadas exclusivamente à classe AFL – apresentou

resultados muito próximos do cenário Original, confirmando a utilidade da *augmentation* para reforçar o equilíbrio entre classes sem distorcer a morfologia do sinal.

No que respeita à interpretabilidade, os métodos de XAI aplicados – SHAP e LIME – permitiram analisar a relevância das diferentes derivações e segmentos temporais do ECG na decisão do modelo. A análise SHAP revelou que as derivações II, III, aVF e V1 apresentaram maior impacto na classificação de AFL, em concordância com a literatura clínica (Cosio 2017), que identifica essas derivações como as mais representativas do padrão em "dente de serra" característico da arritmia. O método LIME, por sua vez, forneceu uma explicação local, destacando os segmentos temporais mais relevantes em exemplos individuais. Embora os resultados do LIME tenham sido, em geral, coerentes com o SHAP, observou-se a atribuição de importância a algumas regiões sem relevância fisiológica direta, o que reflete as limitações inerentes à sua natureza perturbativa, também referidas por Singh e Sharma (2022). Esta divergência reforça a importância de utilizar métodos complementares de XAI para garantir uma explicabilidade mais consistente.

A comparação entre abordagens evidencia ainda que o uso direto de sinais brutos de ECG, aliado a modelos *end-to-end*, oferece uma representação mais completa das dependências temporais e espaciais, dispensando etapas de extração manual de *features*. Contudo, abordagens clássicas baseadas em características extraídas - como intervalos RR, morfologia QRS ou transformadas no domínio da frequência - continuam a ser relevantes pela sua interpretabilidade e eficiência computacional (Acharya et al. 2017; Faust et al. 2018). De forma semelhante, métodos que transformam o ECG em representações bidimensionais (como escalogramas ou espectrogramas) oferecem uma perspetiva tempo-frequência mais rica, conforme explorado por Zeng et al. (2023), embora impliquem maior custo computacional. Assim, conclui-se que a combinação de modelos *end-to-end* com técnicas de extração de *features* ou representações visuais pode potenciar o equilíbrio entre precisão e interpretabilidade, configurando uma linha de investigação promissora.

De forma geral, os resultados obtidos demonstram que as redes CNN–LSTM, complementadas por técnicas de XAI, são capazes de atingir desempenho competitivo em tarefas de classificação de arritmias, mantendo coerência com a fisiologia subjacente dos sinais.

A utilização de métodos explicáveis não só contribui para a transparência do processo de decisão, como também aumenta a confiança dos profissionais de saúde na adoção de sistemas automáticos de apoio à decisão clínica.

5.2 Conclusões

O presente trabalho teve como principal objetivo o desenvolvimento e avaliação de um modelo híbrido CNN–LSTM para a classificação automática de sinais ECG de 12 derivações, recorrendo a métodos de XAI para a interpretação dos resultados. O modelo alcançou valores de desempenho consistentes e robustos, confirmando a eficácia da abordagem *end-to-end* na deteção de padrões eletrofisiológicos complexos.

A aplicação das técnicas SHAP e LIME permitiu identificar as derivações e os segmentos temporais mais relevantes para a decisão do modelo, confirmando a sua coerência com a literatura clínica e demonstrando o potencial das metodologias de explicabilidade em cardiologia computacional. Estes resultados evidenciam que a integração de métodos de XAI é essencial para aumentar a transparência e a confiança em sistemas de inteligência artificial aplicados à saúde.

Entre as principais limitações do trabalho destacam-se o uso de bases de dados públicas e desbalanceadas, a análise a três classes de ritmos cardíacos e a ausência de metadados clínicos (idade, sexo, diagnóstico associado), que poderiam enriquecer a capacidade de generalização do modelo.

Em síntese, este estudo demonstra a aplicabilidade de modelos CNN–LSTM na classificação automática de arritmias cardíacas e reforça a importância da explicabilidade no desenvolvimento de sistemas de apoio à decisão clínica. Os resultados obtidos potenciam a adoção de soluções de inteligência artificial mais transparentes, fiáveis e alinhadas com as exigências éticas e científicas da prática médica moderna.

Como perspetivas futuras, propõe-se a expansão do número de classes consideradas, a integração de metadados clínicos no processo de treino, a utilização de arquiteturas mais avançadas (como *Transformers* com mecanismos de *attention*) e a exploração de novas

técnicas XAI para explicações mais fiáveis e clinicamente interpretáveis.

Referências Bibliográficas

Abdullah, T. A., M. S. M. Zahid, W. Ali e S. U. Hassan (2023). "B-LIME: An improvement of LIME for interpretable deep learning classification of cardiac arrhythmia from ECG signals". Em: *Processes* 11.2, p. 595. DOI: 10.3390/pr11020595. URL: https://www.mdpi.com/2227-9717/11/2/595.

Acharya, U. R., S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych e R. San Tan (2017). "A deep convolutional neural network model to classify heartbeats". Em: *Computers in Biology and Medicine* 89, pp. 389–396. DOI: 10.1016/j.compbiomed.2017.08.022. URL: https://www.sciencedirect.com/science/article/pii/S0010482517302810.

Alamatsaz, N., L. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz e F. Nasimi (2024). "A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection". Em: *Biomedical Signal Processing and Control* 90, p. 105884. DOI: 10.1016/j.bspc.2023.105884. URL: https://www.sciencedirect.com/science/article/pii/S1746809423013174.

Anand, A., T. Kadian, M. K. Shetty e A. Gupta (2022). "Explainable AI decision model for ECG data of cardiac disorders". Em: *Biomedical Signal Processing and Control* 75, p. 103584. DOI: 10.1016/j.bspc.2022.103584. URL: https://www.sciencedirect.com/science/article/abs/pii/S1746809422001069.

Apandi, Z. F. M., R. Ikeura e S. Hayakawa (2018). "Arrhythmia detection using MIT-BIH dataset: A review". Em: *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, pp. 1–5. URL: https://ieeexplore.ieee.org/abstract/document/8477620.

Ardeti, V. A., V. R. Kolluru, G. T. Varghese e R. K. Patjoshi (2023). "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to AI-based approaches". Em: *Expert Systems with Applications* 217, p. 119561. DOI: 10.1016/j.eswa.2023.119561. URL: https://www.sciencedirect.com/science/article/pii/S0957417423000623.

Ayano, Y. M., F. Schwenker, B. D. Dufera e T. G. Debelee (2022). "Interpretable Machine Learning Techniques in ECG-Based Heart Disease Classification: A Systematic Review". Em: *Diagnostics* 13.1, p. 111. DOI: 10.3390/diagnostics13010111. URL: https://www.mdpi.com/2075-4418/13/1/111.

Borghi, P. H., R. C. Borges e J. P. Teixeira (2021). "Atrial fibrillation classification based on MLP networks by extracting Jitter and Shimmer parameters". Em: *Procedia Computer Science* 181, pp. 931–939. DOI: 10.1016/j.procs.2021.12.104. URL: https://www.sciencedirect.com/science/article/pii/S1877050921002921.

Chaddad, A., J. Peng, J. Xu e A. Bouridane (2023). "Survey of Explainable AI Techniques in Healthcare". Em: *Sensors* 23.2, p. 634. DOI: 10.3390/s23020634. URL: https://www.mdpi.com/1424-8220/23/2/634.

Chang, Y. C., S. H. Wu, L. M. Tseng, H. L. Chao e C. H. Ko (2018). "AF detection by exploiting the spectral and temporal characteristics of ECG signals with the LSTM model". Em: *2018 Computing in Cardiology Conference (CinC)*. Vol. 45, pp. 1–4. URL: https://ieeexplore.ieee.org/abstract/document/8743669.

Chen, C., Z. Hua, R. Zhang, G. Liu e W. Wen (2020). "Automated arrhythmia classification based on a combination network of CNN and LSTM". Em: *Biomedical Signal Processing and Control* 57, p. 101819. DOI: 10.1016/j.bspc.2019.101819. URL: https://www.sciencedirect.com/science/article/abs/pii/S1746809419304008.

Cheng, J., Q. Zou e Y. Zhao (2021). "ECG signal classification based on deep CNN and BiLSTM". Em: *BMC Medical Informatics and Decision Making* 21, p. 365. DOI: 10.1186/s12911-021-01691-2.

Cosio, F. G. (2017). "Atrial Flutter, Typical and Atypical: A Review". Em: *Arrhythmia & Electrophysiology Review* 6.2, pp. 55–62. DOI: 10.15420/aer.2017:9:2. URL: https:

//www.aerjournal.com/articles/atrial-flutter-typical-and-atypical-review.

Costa, R., T. Winkert, A. Manhães e J. P. Teixeira (2021). "QRS peaks, P and T waves identification in ECG". Em: *Procedia Computer Science* 181, pp. 957–964. DOI: 10.1016/j.procs.2021.12.107. URL: https://www.sciencedirect.com/science/article/pii/S1877050921002957.

Faust, O., A. Shenfield, M. Kareem, T. R. San, H. Fujita e U. R. Acharya (2018). "Automated Detection Of Atrial Fibrillation Using Long Short-Term Memory Network With RR Interval Signals". Em: *Computers in Biology and Medicine* 102, pp. 327–335. DOI: 10.1016/j.compbiomed.2018.09.020. URL: https://www.sciencedirect.com/science/article/abs/pii/S0010482518301847.

Finocchiaro, G., N. Sheikh, E. Biagini, M. Papadakis, G. Sinagra, A. Pelliccia, C. Rapezzi, S. Sharma e I. Olivotto (2020). "The electrocardiogram in the diagnosis and management of patients with hypertrophic cardiomyopathy". Em: *Heart Rhythm* 17.1, pp. 142–151. DOI: 10.1016/j.hrthm.2019.08.003. URL: https://www.sciencedirect.com/science/article/abs/pii/S1547527119306605.

Ganeshkumar, M., V. Ravi, V. Sowmya, E. A. Gopalakrishnan e K. P. Soman (2023). "Explainable Deep Learning-Based Approach for Multilabel Classification of Electrocardiogram". Em: *IEEE Transactions on Engineering Management* 70.8, pp. 2787–2799. DOI: 10.1109/TEM.2021.3104751. URL: https://ieeexplore.ieee.org/document/9537612.

Ghazarian, A., J. Zheng, D. Struppa e C. Rakovski (2022). "Assessing the reidentification risks posed by deep learning algorithms applied to ECG data". Em: *IEEE Access* 10, pp. 68711–68723. DOI: 10.1109/ACCESS.2022.3187510.

Goodfellow, I., Y. Bengio e A. Courville (2016). *Deep Learning*. MIT Press. URL: https://www.e-hir.org/upload/pdf/hir-22-351.pdf.

Hatamian, F. N., N. Ravikumar, S. Vesal, F. P. Kemeth, M. Struck e A. Maier (2020). "The effect of data augmentation on classification of atrial fibrillation in short single-lead ECG signals using deep neural networks". Em: *ICASSP 2020 - IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1264–1268.

Hochreiter, S. e J. Schmidhuber (1997). "Long short-term memory". Em: *Neural Computation* 9.8, pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Ioffe, S. e C. Szegedy (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift". Em: *arXiv preprint arXiv:1502.03167*. URL: https://arxiv.org/abs/1502.03167.

Karri, M. e C. S. R. Annavarapu (2023). "A real-time embedded system to detect QRS-complex and arrhythmia classification using LSTM through hybridized features". Em: *Expert Systems with Applications* 214, p. 119221. DOI: 10.1016/j.eswa.2022.119221. URL: https://www.sciencedirect.com/science/article/pii/S0957417422022394.

Khurshid, S., S. H. Choi, L.-C. Weng, E. Y. Wang, L. Trinquart, E. J. Benjamin, T. E. Patrick e S. A. Lubitz (2018). "Frequency of cardiac rhythm abnormalities in a half million adults". Em: *Circulation: Arrhythmia and Electrophysiology* 11. DOI: 10.1161/CIRCEP.118.006273. URL: https://www.ahajournals.org/doi/full/10.1161/CIRCEP.118.006273.

Kingma, D. P. e J. Ba (2014). "Adam: A method for stochastic optimization". Em: *arXiv preprint arXiv:1412.6980*. URL: https://arxiv.org/abs/1412.6980.

Kłosowski, G., T. Rymarczyk, D. Wójcik, S. Skowron, T. Cieplak e P. Adamkiewicz (2020). "The Use of Time-Frequency Moments as Inputs of LSTM Network for ECG Signal Classification". Em: *Electronics* 9.9, p. 1452. DOI: 10.3390/electronics9091452. URL: https://www.mdpi.com/2079-9292/9/9/1452.

LeCun, Y., Y. Bengio e G. Hinton (2015). "Deep learning". Em: *Nature* 521.7553, pp. 436–444. DOI: 10.1038/nature14539. URL: https://www.nature.com/articles/nature14539.

Liu, X., H. Wang, Z. Li e L. Qin (2021). "Deep Learning in ECG diagnosis: A review". Em: *Knowledge-Based Systems* 227, p. 107187. DOI: 10.1016/j.knosys.2021.107187.

Lundberg, S. M. e S.-I. Lee (2017). "A Unified Approach to Interpreting Model Predictions". Em: *Advances in Neural Information Processing Systems 30 (NeurIPS)*.

Luz, E. J. D. S., W. R. Schwartz, G. Cámara-Chávez e D. Menotti (2016). "ECG-based heartbeat classification for arrhythmia detection: A survey". Em: *Computer Methods and Programs in Biomedicine* 127, pp. 144–164. DOI: 10.1016/j.cmpb.2015.12.008.

Madan, P., V. Singh, D. P. Singh, M. Diwakar, B. Pant e A. Kishor (2022). "A hybrid deep learning approach for ECG-based arrhythmia classification". Em: *Bioengineering* 9.4, p. 152. DOI: 10.3390/bioengineering9040152. URL: https://www.mdpi.com/2306-5354/9/4/152.

Moningi, R., S. Mahakur, S. Mundada e A. K. Tripathy (2024). "Explainable AI-Based ECG Heartbeat Classification Using Deep Learning Models". Em: *2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–5. URL: https://ieeexplore.ieee.org/abstract/document/10870845.

Nair, V. e G. E. Hinton (2010). "Rectified linear units improve restricted Boltzmann machines". Em: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 807–814.

Nguyen, H. T. T., H. Q. Cao, K. V. T. Nguyen e N. D. K. Pham (2021). "Evaluation of explainable artificial intelligence: SHAP, LIME, and CAM". Em: *Proceedings of the FPT AI Conference*, pp. 1–6. URL: https://www.researchgate.net/publication/362165633.

Organização Mundial de Saúde (2021). *Cardiovascular diseases*. 11 de Junho. URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

Perez Alday, E. A., A. Gu, A. Shah, C. Liu, A. Sharma, S. Seyedi, A. Bahrami Rad, M. Reyna e G. Clifford (2022). *Classification of 12-lead ECGs: The PhysioNet/Computing in Cardiology Challenge 2020 (version 1.0.2)*. PhysioNet. RRID:SCR_007345. DOI: 10.13026/dvyd-kd57. URL: https://doi.org/10.13026/dvyd-kd57.

Perez Alday, E. A., A. Gu, A. J. Shah, C. Robichaux, A. I. Wong, C. Liu, F. Liu, A. B. Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G. D. Clifford e M. A. Reyna (2020). "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020". Em: *Physiological Measurement*. DOI: 10.1088/1361-6579/abc960. URL: http://doi.org/10.1088/1361-6579/abc960.

Ping, Y., C. Chen, L. Wu, Y. Wang e M. Shu (2020). "Automatic detection of atrial fibrillation based on CNN-LSTM and shortcut connection". Em: *Healthcare* 8, p. 139. DOI: 10.3390/healthcare8030139.

Prakash, A. J., K. K. Patro, S. Saunak, P. Sasmal, P. L. Kumari e T. Geetamma (2023). "A new approach of transparent and explainable artificial intelligence technique for patient-specific ECG beat classification". Em: *IEEE Sensors Letters* 7.5, pp. 1–4. DOI: 10.1109/LSENS.2023.3257165. URL: https://ieeexplore.ieee.org/abstract/document/10105971.

Raghu, A., D. Shanmugam, E. Pomerantsev, J. Guttag e C. M. Stultz (2022). "Data augmentation for electrocardiograms". Em: *Conference on Health, Inference, and Learning (CHIL)*. PMLR, pp. 282–310.

Raza, A., K. P. Tran, L. Koehl e S. Li (2022). "Designing ECG monitoring healthcare system with federated transfer learning and explainable AI". Em: *Knowledge-Based Systems* 236, p. 107763. DOI: 10.1016/j.knosys.2021.107763. URL: https://www.sciencedirect.com/science/article/pii/S0950705121011133.

Ribeiro, M. T., S. Singh e C. Guestrin (2016). "Why Should I Trust You? Explaining the Predictions of Any Classifier". Em: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144. DOI: 10.1145/2939672.2939778. URL: https://dl.acm.org/doi/abs/10.1145/2939672.2939778.

Saini, S. K. e R. Gupta (2022). "Artificial Intelligence methods for analysis of electrocardiogram signals for cardiac abnormalities: state-of-the-art and future challenges". Em: *Artificial Intelligence Review* 55, pp. 1519–1565. DOI: 10.1007/s10462-021-09999-7. URL: https://link.springer.com/article/10.1007/s10462-021-09999-7.

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview". Em: *Neural Networks* 61, pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

Senthuran, V., U. Thayasivam, I. Natgunanathan, K. Sood e Y. Xiang (2025). "Balancing privacy and health integrity: A novel framework for ECG signal analysis in immersive environments". Em: *Computers in Biology and Medicine* 192, p. 110234.

Singh, P. e A. Sharma (2022). "Interpretation and classification of arrhythmia using deep convolutional network". Em: *IEEE Transactions on Instrumentation and Measurement* 71, pp. 1–12. DOI: 10.1109/TIM.2022.3202764. URL: https://ieeexplore.ieee.org/abstract/document/9877905.

Sinha, N., R. K. Tripathy e A. Das (2022). "ECG beat classification based on discriminative multilevel feature analysis and deep learning approach". Em: *Biomedical Signal Processing and Control* 78, p. 103943. DOI: 10.1016/j.bspc.2022.103943. URL: https://www.sciencedirect.com/science/article/pii/S1746809422004426.

Sreeja, M. U. e M. H. Supriya (2023). "A Deep Convolutional Model for Heart Disease Prediction based on ECG Data with Explainable AI". Em: *WSEAS Transactions on Information Science and Applications* 20, pp. 254–264. URL: https://wseas.com/journals/isa/2023/a585109-015(2023).pdf.

Teixeira, J. P. e V. Lopes (2011). "Help system for medical diagnosis of the electrocardiogram". Em: *1st International Conference on Serious Games and Applications for Health*. Braga, pp. 95–102. URL: https://bibliotecadigital.ipb.pt/handle/10198/9708.

Wagner, P., N. Strodthoff, R. Bousseljot, W. Samek e T. Schaeffter (2022). *PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3)*. PhysioNet. RRID:SCR_007345.

Wagner, P., N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek e T. Schaeffter (2020). "PTB-XL: A Large Publicly Available ECG Dataset". Em: *Scientific Data*. DOI: 10.1038/s41597-020-0495-6.

Wasimuddin, M., K. Elleithy, A. S. Abuzneid, M. Faezipour e O. Abuzaghleh (2020). "Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey". Em: *IEEE Access* 8, pp. 177782–177803. DOI: 10.1109/ACCESS.2020.3027056. URL: https://ieeexplore.ieee.org/abstract/document/9206538.

Yao, Q., R. Wang, X. Fan, J. Liu e Y. Li (2020). "Multi-class arrhythmia detection from 12-lead varied-length ECG using attention-based time-incremental convolutional

neural network". Em: *Information Fusion* 53, pp. 174–182. DOI: 10.1016/j.inffus.2019.06.024.

Yıldırım, O., P. Pławiak, R. Tan e U. R. Acharya (2018). "Arrhythmia detection using deep convolutional neural network with long duration ECG signals". Em: *Computers in Biology and Medicine* 102, pp. 411–420. DOI: 10.1016/j.compbiomed.2018.09.009. URL: https://www.sciencedirect.com/science/article/abs/pii/S0010482518302713.

Zeng, W., L. Shan, C. Yuan e S. Du (2024). "Advancing cardiac diagnostics: Exceptional accuracy in abnormal ECG signal classification with cascading deep learning and explainability analysis". Em: *Applied Soft Computing* 165, p. 112056. DOI: 10.1016/j.asoc.2024.112056. URL: https://www.sciencedirect.com/science/article/abs/pii/S1568494624008305.

Zeng, W., B. Su, Y. Chen e C. Yuan (2023). "Arrhythmia Detection Using TQWT, CEEMD and Deep CNN-LSTM Neural Networks with ECG Signals". Em: *Multimedia Tools and Applications* 82.19, pp. 29913–29941. DOI: 10.1007/s11042-023-15655-0. URL: https://link.springer.com/article/10.1007/s11042-023-15655-0.

Zheng, J., H. Chu, D. Struppa, J. Zhang, S. M. Yacoub, H. El-Askary, A. Chang, L. Ehwerhemuepha, I. Abudayyeh, A. S. Barrett, G. Fu, H. Yao, D. Li, H. Guo e C. Rakovski (2020). "Optimal Multi-Stage Arrhythmia Classification Approach". Em: *Scientific Reports* 10. DOI: 10.1038/s41598-020-73581-7.

Zheng, J., H. Guo e H. Chu (2022). *A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0)*. PhysioNet. RRID:SCR_007345.